TEC-0106

Research in Model-Based Change Detection and Site Model Updating

R. Nevatia
A. Huertas

University of Southern California
Institute for Robotics and Intelligent Systems
Powell Hall 204
Los Angeles, CA 90089-0273

March 1998

Prepared for:
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington, VA 22203-1714

Monitored by:
U.S. Army Corps of Engineers
Topographic Engineering Center
7701 Telegraph Road
Alexandria, VA 22315-3864

US Army Corps of Engineers
Topographic Engineering Center

T
E
C

Destroy this report when no longer needed.
Do not return it to the originator.

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

The citation in this report of trade names of commercially available products does not constitute official endorsement or approval of the use of such products.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	March 1998	Final July 1993 - July 1996

4. TITLE AND SUBTITLE

Research in Model-Based Change Detection and Site Model Updating

5. FUNDING NUMBERS

DACA76-93-C-0014

6. AUTHOR(S)

R. Nevatia A. Huertas

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Southern California
Institute for Robotics and Intelligent Systems
Powell Hall 204
Los Angeles, CA 90089-0273

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Defense Advanced Research Projects Agency
3701 North Fairfax Drive, Arlington, VA 22203-1714

U.S. Army Topographic Engineering Center
7701 Telegraph Road, Alexandria, VA 22315-3864

19. SPONSORING / MONITORING AGENCY REPORT NUMBER

TEC-0106

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

USC's major research project deals with change detection and site model updating. A system was developed to validate building models and detect changes, i.e. in the dimensions of a building. The system uses the site model to predict the location and position of structures in the image. The comparison between model and image features gives an indication of the changes to the structures. Updating the model by adding descriptions of new structures requires site construction techniques. CMU has developed two systems for site model construction with focus on building structures. These systems implement perceptual grouping techniques to detect and describe the buildings in the scene by means of 3-D models. One system uses a single image to derive a model with the aid of 3-D clues, such as shadows and visible walls. Results from more than one image are combined to reinforce detection confidence and attempt to complete partial detection. This system also incorporates facilities for user interaction that help construct high quality site models with minimal interaction. The second system uses multiple images at all levels of processing and uses shadows and visible wall information to hypothesize and verify buildings and construct 3-D models. Results from RADIUS modelboard are presented.

14. SUBJECT TERMS

Change Detection Model Construction Model Updating

15. NUMBER OF PAGES

75

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

TABLE OF CONTENTS

ILLUSTRATIONS

FIGURE PAGE

TABLES

TABLE TITLE **PAGE**

Preface

This research was sponsored by the Defence Advanced Research Projects Agency (DARPA) and monitored by the U.S. Army Topographic Engineering Center (TEC) under contract DACA76-93-C-0014, titled "Research in Model-Based Change Detection and Model Updating." The DARPA Program Managers were Dr. Oscar Firschein and Dr. Tom Strat, and the TEC Contracting Officer's Representative was Ms. Lauretta Williams.

1 Introduction

This report describes our activities during the period of 28 July, 1995 to 27 July, 1996, the third and final year of this effort. Along with the previous annual reports, this constitutes our "final technical report." The primary focus of this work has been on change detection and site model updating. Methods have been developed for detecting changes to fixed structures, such as buildings, and to detect presence or absence of large mobile objects, such as aircraft. We have continued to develop automated and methods for building detection and description, using either monocular or multiple images. These techniques are needed for automated site model construction and model updating. We have also developed a method for interacting with the automatic site modeling system that requires minimal interaction from a human user. These projects are briefly described below in this section, and details are given in the following sections.

1.1 Change Detection

The task of change detection consists of finding significant differences between the new data and the model derived from the older data. The significance of the differences may be task specific though in most cases man-made changes are more important than those caused by natural factors such as seasonal changes. We are only interested in those changes in the image that come from some changes in the site rather than from changes in imaging conditions.

The first step in change detection, *Site Model to Image Registration*, is to register the new image(s) to the model(s) contained in the site folder. The system has some capability for performing coarse registration, however, this information is expected to be available from other modules being developed by other contractors under the RADIUS program. The system uses feature matching [1] to compensate globally for traslational errors and bring the site model into close correspondence with the observed image.

The second step, *Site Model Validation*, verifies the presence of the model objects in the image. A confidence value is computed for each object in the model based on the match information from the previous step. Lower confidence values are likely to represent possible changes to the objects.

In the third step, *Change Detection*, possible change in the site indicated in the previous step is analyzed in more detail, and it is determined if the missing correspondences can be explained by techniques that draw attention to significant structures in the image that are not explained by the existing model. The task of finding objects in the image that are not in the model is more difficult, and will require use of *perceptual grouping* operations. Currently the system can detect missing buildings and changes in dimensions.

In the fourth step, *Detailed Analysis*, the structures indicated by the change detection processes are analyzed in detail. This step requires the development of automated or semi-automated site modeling techniques.

In the fifth step, *Site Model Updating*, the changes are modeled and incorporated in the new site model.

Our previous annual reports [2, 3] and [4] gave details on the development of a validation system that included fine registration followed by a simple object-by-object verification scheme. The scheme only measured validation by counting the number of object elements matched to image features. During the past year we have made several improvements [5], and have continued testing the model registration and validation system with emphasis on detection of changes to the structures in the model.

We have made improvements in the validation technique in two ways: We incorporated the use of junctions into the validation scores. The validation scores give a measure of belief on the evidence that supports the continued presence of an object in the scene. Previously we had considered edge support only, and by adding the additional evidence of expected junctions we obtain a more robust measure. Second, we implemented the procedures needed to validate and analyze possible changes for all the model buildings in a site. This addition renders the validation/change detection process more useful in practical terms as it permits the processing of an entire image regardless of its size. This new process, however, assumes that the camera models associated with the various site views is accurate enough (within three pixels) to allow processing small portions of the image surrounding each object. Thus processing time becomes a function of the number of objects to be validated and checked for changes. Grossly misplaced individual structures are then reported as missing in the image or as incorrectly placed in the model.

The validation and change detection system, which uses the fast block interpolation projection (FBIP) camera model, is written in LISP and runs under the Radius Common Development Environment (RCDE) [6]. It has been integrated in the RADIUS testbed system [7] developed by the Lockheed-Martin Corp., and delivered to the NEL for further testing and evaluation.

The validation system also has been applied to the task of verifying the presence of aircraft at a site. The suggested method involves the derivation of simple aircraft models from one or more images, rather than using CAD models. These results assume that the pose of the aircraft and the sun angles are known *a priori*. Details are given in [8].

1.2 Automated Building Detection and Description

We have continued the work in automatic building detection and description. This ability is needed for reliable change detection and site model updating, and is also useful for initial site model construction. Two systems are under development: one uses a single intensity image and another uses multiple images. It is, of course, easier to detect and describe buildings using multiple images, however, the ability to at least reliably detect buildings from a single image is needed during the change detection process.

Good progress has been made on both systems. The monocular system [9] [10] now uses both shadows and walls for verification of a building, and for estimating heights. It has been tested extensively on the modelboard images with good results. These are presented in Section 3. More recently, we have tested the system with Fort Hood images with very good results as well [11]. Currently, this system is used to detect building structures not present in the site model, to update the model. This system also has been integrated into the RADIUS testbed system.

The system using multiple images is in earlier stages of development. The system uses a hierarchical grouping and matching methodology [12]. The preliminary results are encouraging and we believe that this method will lead to robust and reliable building detection and description. Software has been delivered to the Lockheed-Martin Corp., for testing. This system is described in Section 4.

1.3 Interaction with Automatic Model Construction Systems

Another area of progress, described in Section 3, deals with user interaction with the automated systems to assist the building detection systems in completion of the modeling task [13]. The general idea consists of identifying areas, or cases, where the automated systems fail. The user guides the automated systems to use the partial results derived automatically to help complete the task with minimal user input. This system has been tested in conjunction with our monocular building detection system with encouraging results.

Acknowledgments

The work reported here is the result of contributions from several faculty, research staff, graduate students and visiting researchers, under the direction of Prof. Ram Nevatia. The principal developers were Chungan Lin (monocular model construction), Sanjai Noronha (multi-view model construction) and Andres Huertas (change detection). Mathias Bejanin (model registration), Stephane Maruani (aircraft validation) and Stephan Heuel (user interaction for model construction) were visitors to USC. Prof. Keith Price assisted in all aspects related to the use of the RCDE and LISP programming in general. Profs. Nevatia and Price carried out the testing of the USC systems using operational imagery, and coordinated the technology transfer efforts to SRI international, the Lockheed-Martin Corp., and the NEL.

2 Change Detection and Model Updating

Change detection is an important task in the process of photo-interpretation. It also is a tedious task as it requires careful comparison of images, and their models, taken at different times under possibly varying conditions. We believe that even partial automation of this task will greatly increase image analyst productivity and also possibly enhance the reliability of the results.

The task of change detection consists of finding significant differences between the new data and the model derived from the older data. The significance of the differences may be task specific though, in most cases, man-made changes are more important than those caused by natural factors such as seasonal changes. We are only interested in those changes in the image that come from some changes in the site rather than from changes in imaging conditions.

Change detection involves comparing a *new* image (or a collection of images) of a site to the information associated with that site in a *site folder.* This information may consist of a site model and one or more previous images, and results of previous analyses on these images. We assume that in all cases a site model of suitable resolution and complexity is available.

Figure 2.1 shows a flowchart of the complete change detection system. It contains five major steps:

- *Site Model to Image Registration*: The first step in change detection is to register the new image(s) to the model(s) contained in the site folder. The system uses feature matching [1] to compensate globally for translational errors and bring the site model into close correspondence with the observed image.
- *Site Model Validation*: This step verifies the presence of the model objects in the image. A confidence value is computed for each object in the model based on the match information from the previous step. Lower confidence values are likely to represent possible changes to the objects.
- *Change Detection:* In this step we analyze, in more detail, possible change in the site indicated in the previous step, and determine if the missing correspondences can be explained by techniques that draw attention to significant structures in the image that are not explained by the existing model. The task of finding objects in the image that are not in the model is more difficult, and will require use of *perceptual grouping* operations [9,10,11,12,15, 16]. Currently the system can detect missing buildings, changes in dimensions, and new buildings under some conditions.
- *Detailed Analysis:* In this step, the structures indicated by the change detection processes are analyzed in detail. This step may require the use of more than one view of the scene. This step also requires the development of automated or semi-automated site modeling techniques
- *Site Model Updating*: In this step the changes are modeled and incorporated in the new site model.

In the following we describe our work on tasks needed to achieve a full change detection system: *site model registration and validation, and preliminary change detection.* We show examples that illustrate the major steps and summarize results of experiments using site models and associated imagery supplied to us

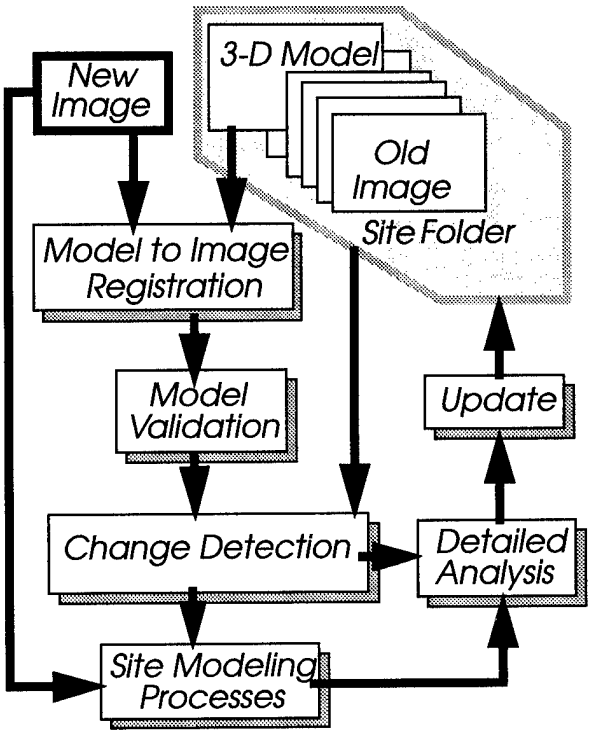


Figure 2.1 Flowchart of the change detection system

by the RADIUS program [7]. This system uses the fast block interpolation projection (FBIP) camera model, is written in LISP, and runs under the RCDE [6].

2.1 Site Model to Image Registration

The first step is to register the site model to an image. Normally, coarse registration should be available from other modules of the RADIUS program. Our system has capability to correct for translational errors. Our registration method [2, 4] consists of the following tasks:

- Calculation of misregistration offsets and compensation for translational errors.
- Establishment of correspondences between the elements of objects in the model, and the supporting features extracted from the image.

The first task is carried out by a matching technique [1] that uses line segments derived from the site model objects, and line segments [17] approximated from the edges extracted [18] from the image. The second task uses the registration offsets to select the matching pairs (model segment, image segment) that correspond the model objects to the image features.

Figure 2.2 shows an example of the registration step in the system. The site model shown in Figure 2.2a is projected according to the camera model associated with the image. The peak in the matcher accumulator array (Figure 2.2b) gives the global misregistration error. The fine-registered model (Figure 2.2c) is then used to establish the context needed for further processing. Details may be found in [2] and [4].

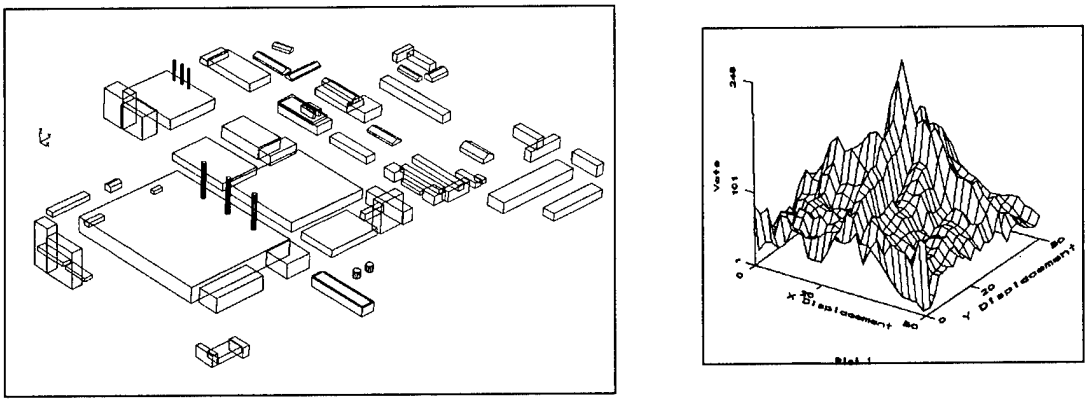


Figure 2.2 (a) Site Model. (b) Accumulator array

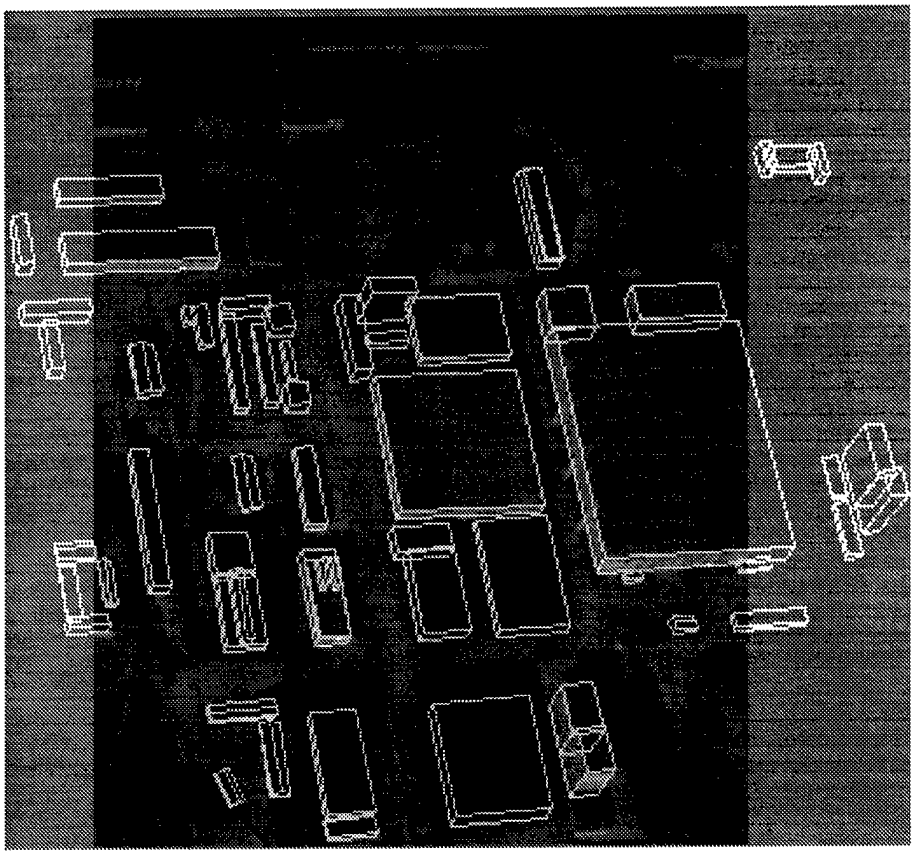


Figure 2.2 (continued)(c) Site model registered with image

2.2 Site Model Validation

The purpose of model validation is to verify that model objects are present in the image. The system uses the correspondences established in the registration step to assign a confidence value to each object in the model to reflect its image support, and to help select object candidates to analyze for likely changes in the site. Some missing features will be caused by viewing conditions. These, however, can be predicted

and explained from the site model itself. The system, at this stage, also deals with ambiguities, such as multiple matches and coincidental alignments.

The confidence values derived are based on the following measures:

Object Presence: Each object model consists of a number of edges representing its boundaries. Object presence denotes how many of these boundaries have a corresponding segment or segments in the image (see Figure 2.3) Currently, object presence is measured as a percentage of model edges corresponded to image edges. This quantity takes into account only visible elements, from the particular viewpoint of the image. Both self-occlusion and occlusion by other objects are determined using the range image derived from the model itself. Thus non-visible objects are not counted.

Object presence is also calculated separately for **roof** elements, **vertical** elements and **base** elements to allow us to study the relative importance of these components as a function of the viewpoint. Near-nadir views, for example, may highlight the contribution of the roof elements. These weights may be set by annotations in the site model; currently, they are given equal weight.

Object Coverage: Object coverage is equivalent to length-weighted object presence. It denotes the percentage of the perimeters of the matched boundaries of the faces of the objects. These quantities represent the amount of boundary evidence detected in the image in support of the validation of a model object. Figure 2.3a shows an object with all sides represented by small supports. Figure 2.3b shows the opposite; a few sides represented with good support. Object coverage measurements take into account occlusion and are calculated separately for roof, vertical and base elements.

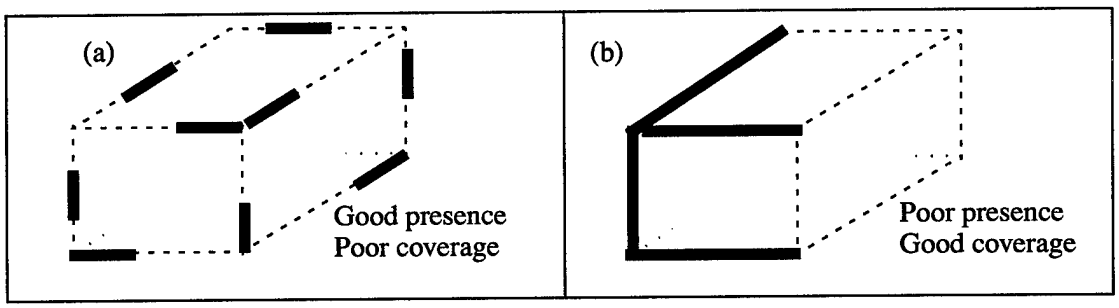


Figure 2.3 Presence and Coverage

Shadow Presence: Shadow presence is inferred from the models and verified in the image. The model information is used to project the shadow boundaries, taking into account their visibility. Note that in situations where reasonable object matches (correspondences) are not found, the absence of shadows help confirm the absence of the building, but the presence of shadows does not guarantee the presence of a building. The final interpretation is the subject of our current and future work.

The measure for shadow presence is defined as the ratio of the number of potential shadow boundaries and junctions extracted from the image, over the number of visible shadow elements (boundaries and junctions) derived from the model (see Figure 2.4). The image segments are labelled as potential shadow segments by noting the consistency of the "dark" side of the segment with respect to the direction of illumination. Segments oriented parallel to the direction of illumination also correspond to possible shadow lines cast by vertical object edges. Shadow junctions are detected similarly. The L-junctions formed (allowing for gaps) by potential shadow lines are labeled potential shadow junctions. Details on the shadow labeling of segments and junctions may be found in [10] and [11].

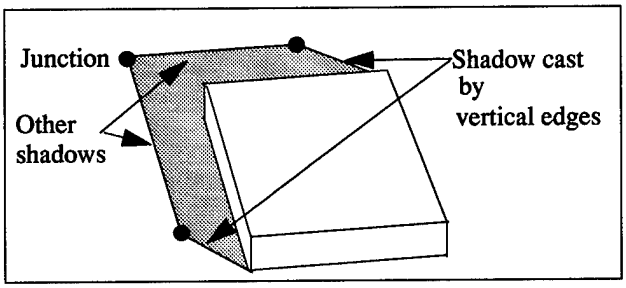


Figure 2.4 Shadows cast by "cubic" building

Object presence and coverage and shadow presence are currently combined to give a match value from which a confidence level is established (see Table 1), as follows:

$$\text{Match Value} = \frac{1}{\sum_i w_i} \cdot \left[w_p \cdot P(x) + w_v \cdot V(x) + w_s \cdot S(x) + w_j \cdot J(x) + w_m \cdot \frac{M(x)}{F(x)} \right]$$

Where: $P(x)$ measures "presence," that is, whether an object element is represented in the image; $V(x)$ measures visibility, the fraction of object elements in the field of view and not occluded; $S(x)$ measures the presence of shadows, that is, whether or not shadow elements predicted from the model are represented in the image; $M(x)$ measures "coverage", the portion of the visible model elements matched to image elements; $J(x)$ measures the presence of matching junctions, that is, whether or not junctions between model elements are found to be present in the image; and $F(x)$ measures boundary fragmentation, in terms the number of image elements matched to model elements. The weight assignment is arbitrary.

Table 1. Confidence Levels

Match Value	**> 0.7**	**> 0.5**	**0.4 - 0.5**	**< 0.4**	**< 0.2**
Confidence	Very High	High	Medium	Low	Very Low
Color	Green	Cyan	Yellow	Salmon	Red

High match values indicate good image support while low values denote low image support. Low values may signify change as lack of image support may be due to missing buildings, or buildings that have undergone significant change with respect to their current model. Model buildings that have high match values, that is, strong image support, may have changed also. Additions to structures, such as new wings, may not significantly affect the appearance of the previously modeled portions. Examples of these situations are shown in the results section. Figure 2.5 shows an example of the registration/validation step applied to one of the modelboard 1 images where there are no changes. The colors indicate the confidence level associated with each building structure. Cyan and green colors represent high values. Yellow represents medium values and orange and red represent low values.

2.3 Change Detection

The confidence values computed in the previous step give the first indication, for each object, of potential changes in the site. High values indicate close correspondence between model and image. Low values signify possible changes to the site. In some cases, however, high values are due to multiple matches

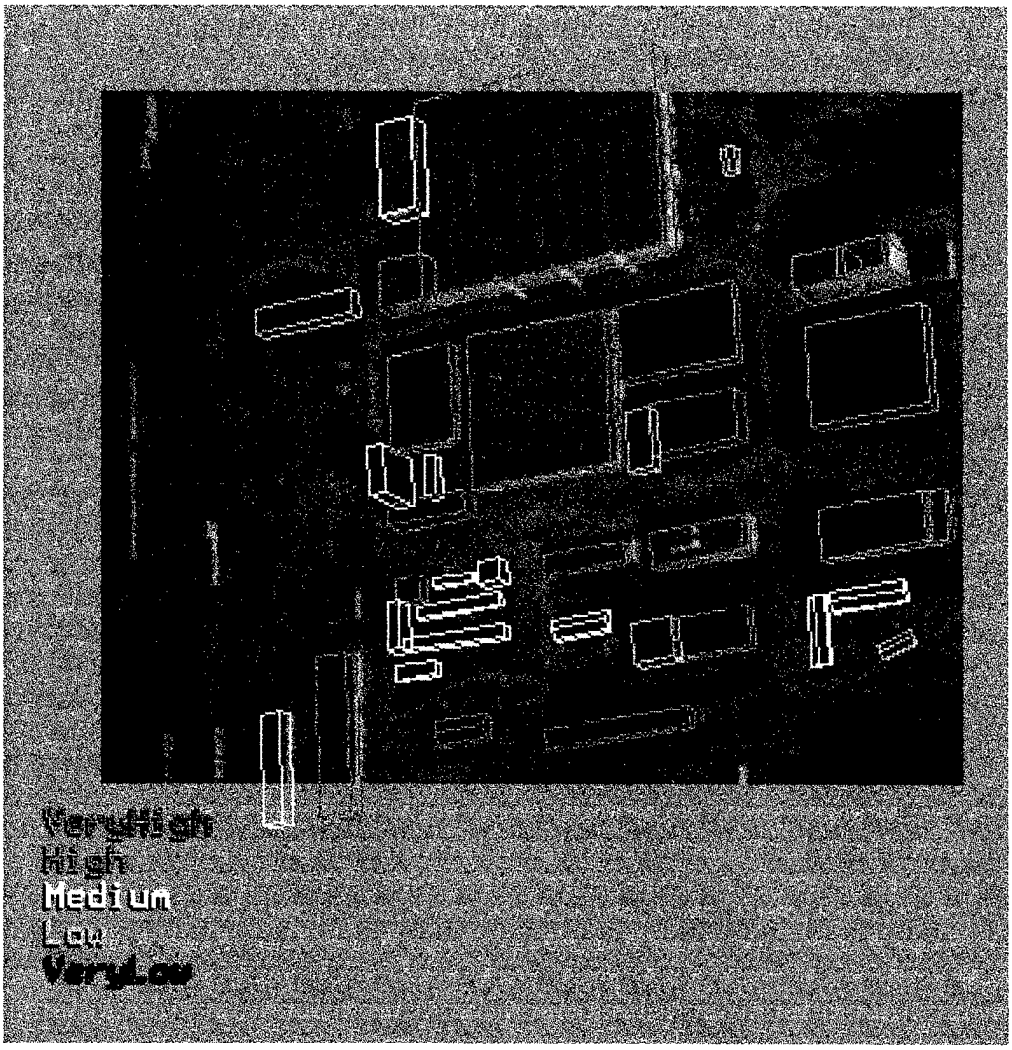


Figure 2.5 Validation result and color-coded confidence levels

and other ambiguities that may exaggerate or reduce image support for an object. These conditions, however, are isolated. To distinguish between apparent and actual changes we first perform an analysis of possible ambiguities and correct the confidence values appropriately, as discussed in the following.

2.3.1 Analysis of Ambiguities

There are two kinds of ambiguities that are resolved by the system. The first deals with multiple or missing matches between the site model features and the image. The second deals with coincidental alignments caused by the viewpoint or to the adjacency of the structures.

2.3.2 Multiple and Missing Matches

The model-to-image matcher in the system corresponds each model element with one or more image elements. This is necessary to deal with expected fragmentation in the image elements. Fragmentation is caused by inadequacies in the feature extraction process, and by actual image content, such as trees occluding buildings or by road boundaries and shadows. This may result in model segments being corresponded to multiple image boundaries (Figure 2.6) or to boundaries of other nearby objects. This condition is detected by observing the object coverage measures described above and is handled in the following manner:

If the multiple matches include colinear image segments, these are currently taken together. If the multiple matches involve parallel image segments, the one with the closest fit to the model segment is taken to represent the matched boundary (see also example below.)

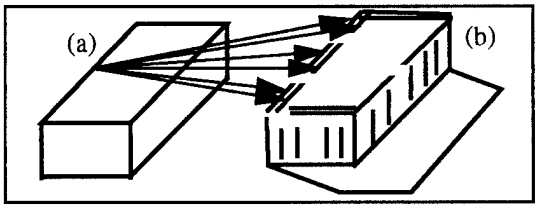


Figure 2.6 Model to image correspondence

In some cases complex objects are overmodeled, i.e. they are modeled in terms of shapes that may include some elements that do not correspond to actual physical elements or boundaries. Figure 2.7 shows an L-shaped building that has been modeled by two rectangle parallelepipeds. The thick lines represent portions of the elements on the building model that do not correspond to physical boundaries. These cannot be matched and are missing. The reduced image support results in lower confidence.

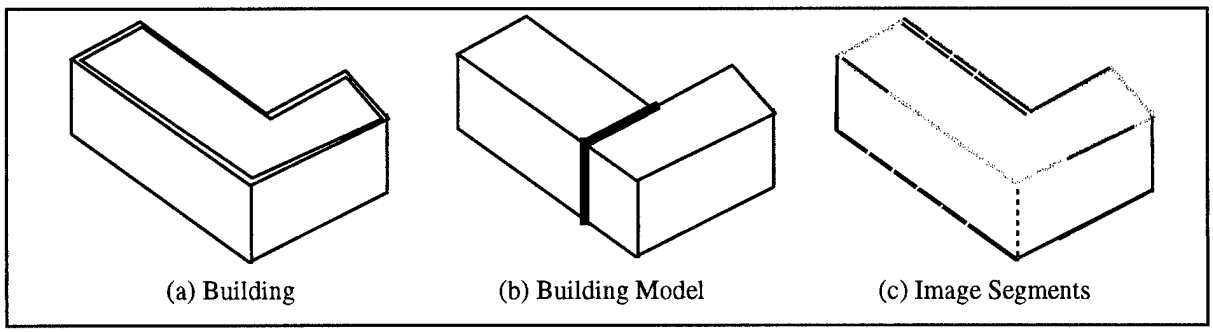


Figure 2.7 Missing match due to overmodeling

Figure 2.8 shows two buildings that are likely to be undermodeled (i.e. modeled by simpler shapes) because of their complexity. These are likely to require additional search strategies designed to look for additional evidence, such as a large number of vertical or horizontal boundaries. The system is not currently capable of determining these conditions, and thus, the confidence values may be underestimated. It is assumed that some of these conditions may require annotations in the site model to help the system adjust the weights used to determine confidence values.

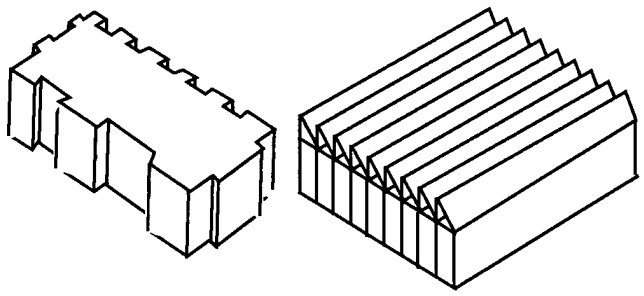


Figure 2.8 Complex buildings may be undermodeled

Next, an example from the modelboard is used to illustrate our previous discussion, and help explain the remaining conditions that the system can currently handle. Figure 2.9(a) shows the model segments with the elements that might have changed as thick lines. A number of possible changes are denoted by circles on the structures. The corresponding image segments are also shown in Figure 2.9(b) as thick lines. In Figure 2.11, the thick black and white lines denote ambiguous multiple matches. After resolution of the ambiguity, the white lines denote the ones chosen to correspond to model edges.

2.3.2.1 Coincidental Alignments

Some of the multiple matches described in the previous section are due to coincidental alignments of buildings with other structures. Some of these include roads, and adjacent objects. Nearby objects and shadows sometimes result in image features that have a larger extent than that predicted by the model features. These are explained by examining nearby shadows with knowledge of the direction of illumination, and by examining adjacent structures.

The building on the top right of Figure 2.11 has a vertical edge aligned with the shadow cast by the same edge. Both edges in the image, the vertical edges and its shadow, are good candidates to match the model's vertical edge. The multiple match may indicate an increase in height but, in this case, the situation is identified correctly as a coincidental alignment. The white portion of the edge is then determined to be the portion corresponding to the model edge.

Coincidental alignments caused by nearby and adjacent structures are determined by locating adjacent structures that help explain a possible change. The small building on the top of Figure 2.11 helps to illustrate this point. The model roof and base edges are matched to much longer lines in the image. Figure 2.10 shows two buildings (white boundaries) that were found to explain the situation detected, thus dismissing the possibility of determining a change in the horizontal dimensions of the small buildings (black boundaries).

In this example, all possible changes are explained by resolving ambiguities in the matching process, and by detecting coincidental alignments with shadows or nearby structures. Therefore no changes are reported.

2.3.3 Changes in the Site

Our system currently is able to detect changes in the dimensions of the structures and changes due to missing buildings. In the following examples we altered the site model to test these conditions.

2.3.3.1 Changed Objects

Changes in the dimensions of the structures located in the image that are not due to errors or coincidental alignment signify real change. The changes in dimensions detected by the current system are preliminary in the sense that they are not fully described. A final determination of change requires that the entire object geometry be analyzed for consistency in view of the possible change. Also, this process may require using more than one view. This is a subject of our future work.

Figure 2.12 shows an example, also from the modelboard image set. The models of the two buildings were altered by hand (reduced in size) to the dimensions illustrated by the thin white lines. The matching and fine registration step correctly registers the modified models to the structures in the image. The thick white lines are the image segments that matched the corresponding model edges. The differences then denote the extent of the change found at this preliminary stage.

Figure 2.13 shows a building wing that has been added to an existing structure. The portion of the building in the model is correctly registered to the image. The two thick white lines denote the extent of

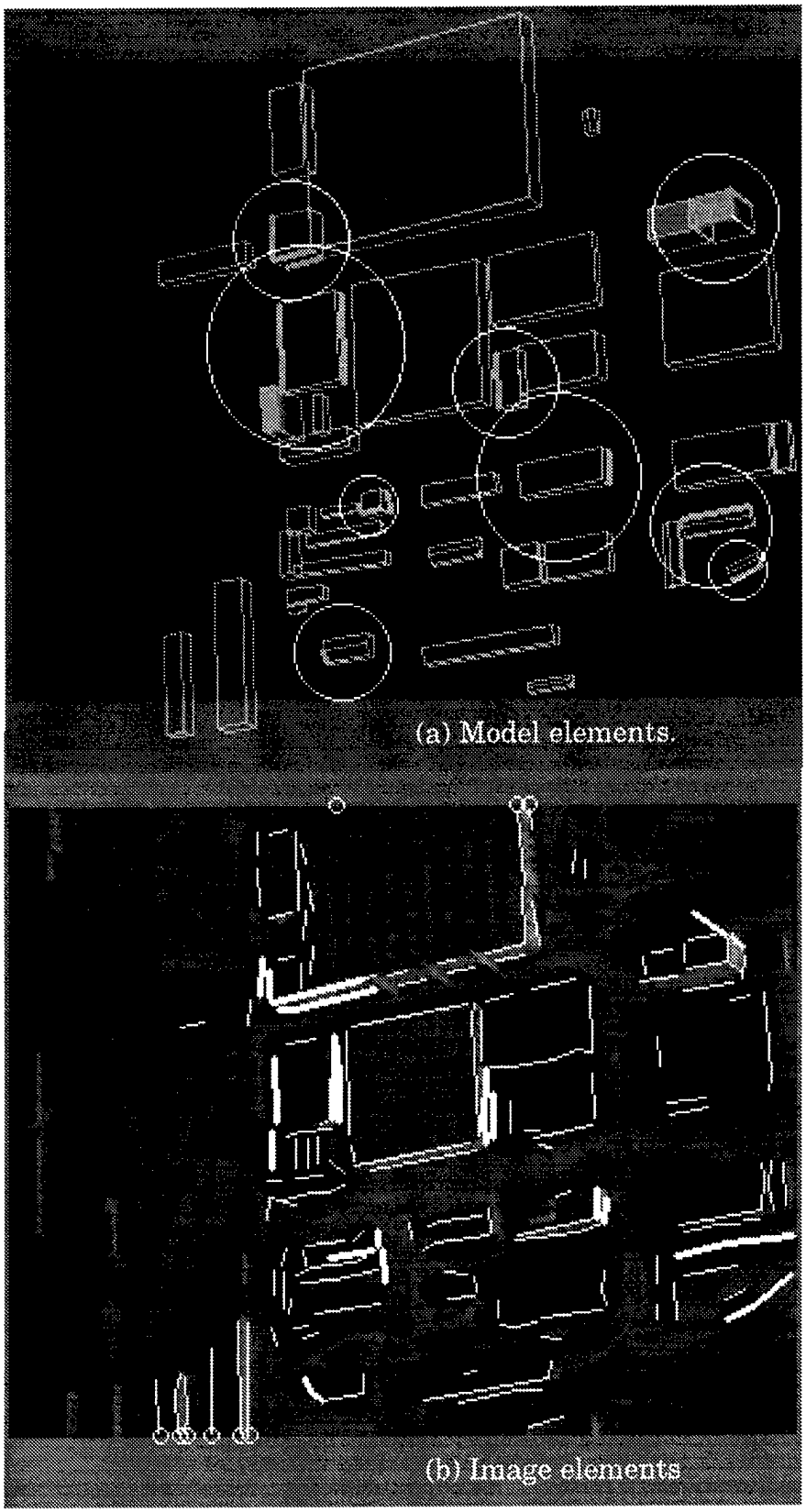


Figure 2.9 Possible changes to be explored at circled locations

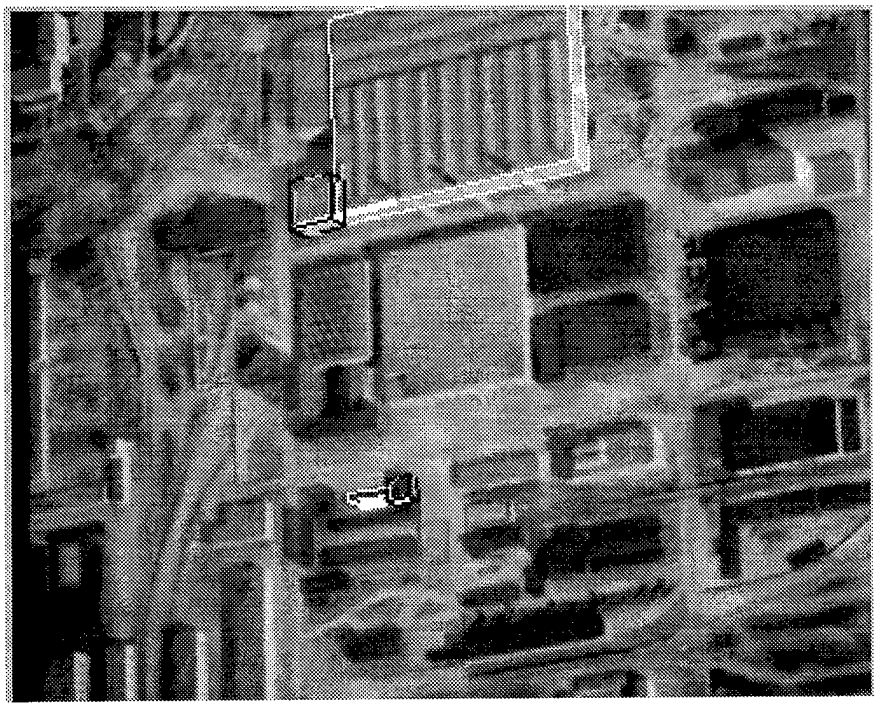


Figure 2.10 Adjacent buildings may introduce ambiguity

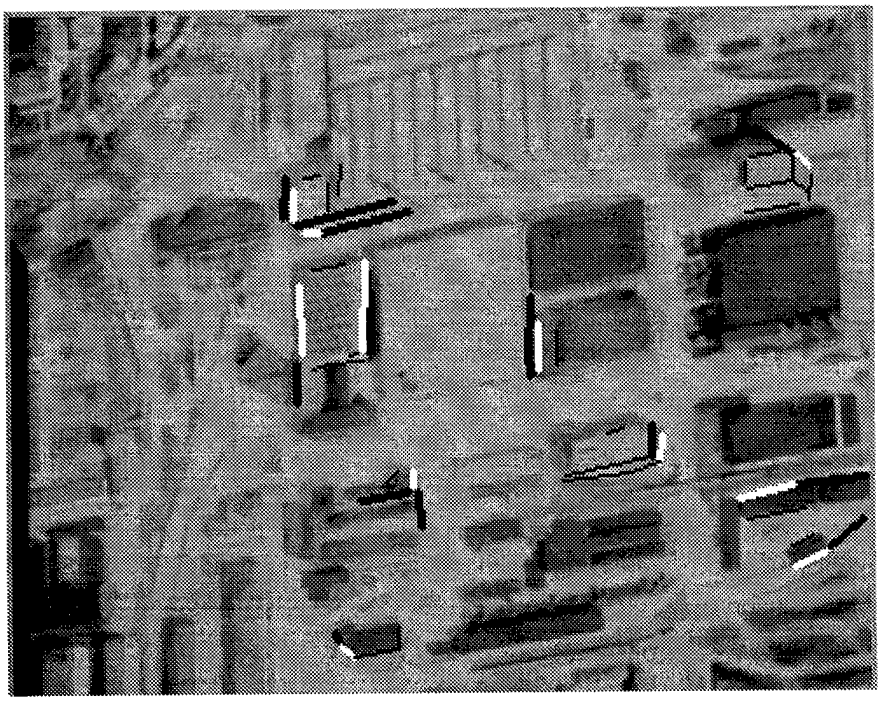


Figure 2.11 Ambiguity due to multiple matches and alignment

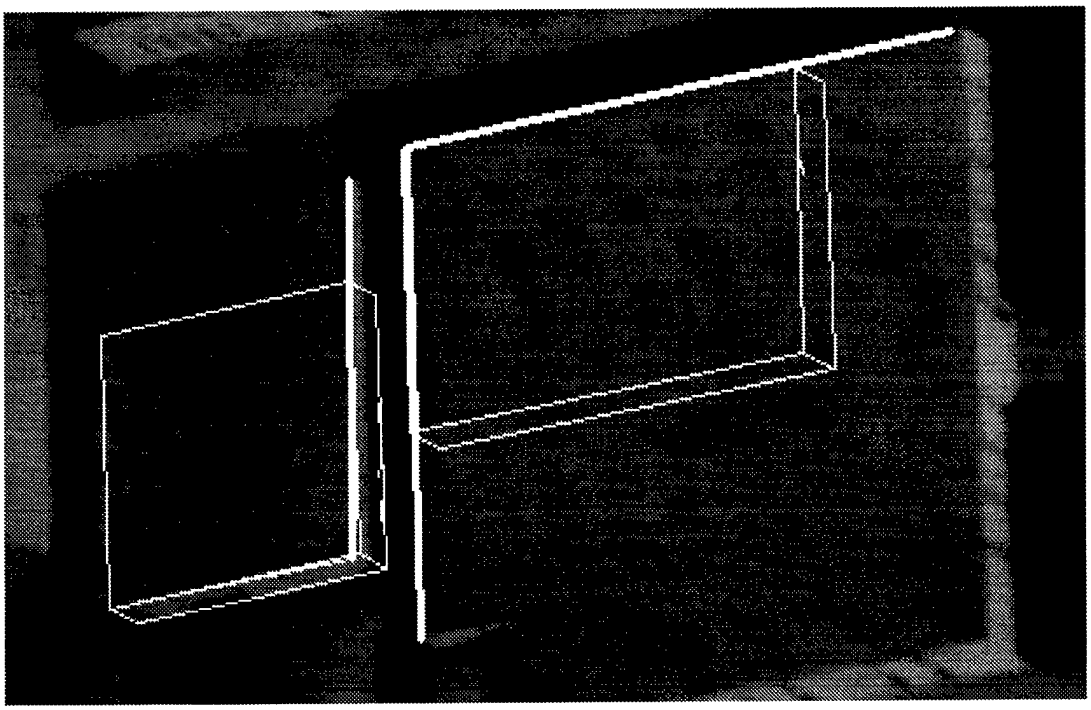


Figure 2.12 Actual change in dimensions

the match. Because the object presence measure for the roof of this structure indicates that all four sides of the current model were matched, the change is labeled "added" wing.

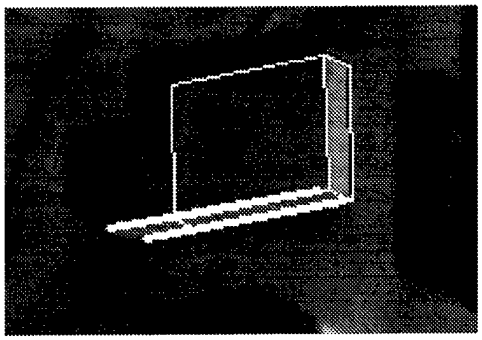


Figure 2.13 Added "wing" is reported in this case

2.3.3.2 Missing Buildings

Figure 2.14 shows a large number of object models (in white) added by hand to the site model. The size and location of these objects were determined randomly and added deliberately to the site model to test for "missing" object capability. Note that, in spite of the added information, the "legitimate" models are correctly registered with the image, as shown by the black lines. The low confidence values calculated indicate that there is no image evidence to support the presence of a building at that location. The two possible causes for this condition are that either the model is incorrect or the building has been removed or destroyed (assuming that images are of sufficient quality.) Resolving these ambiguities may require examination of this location in other images.

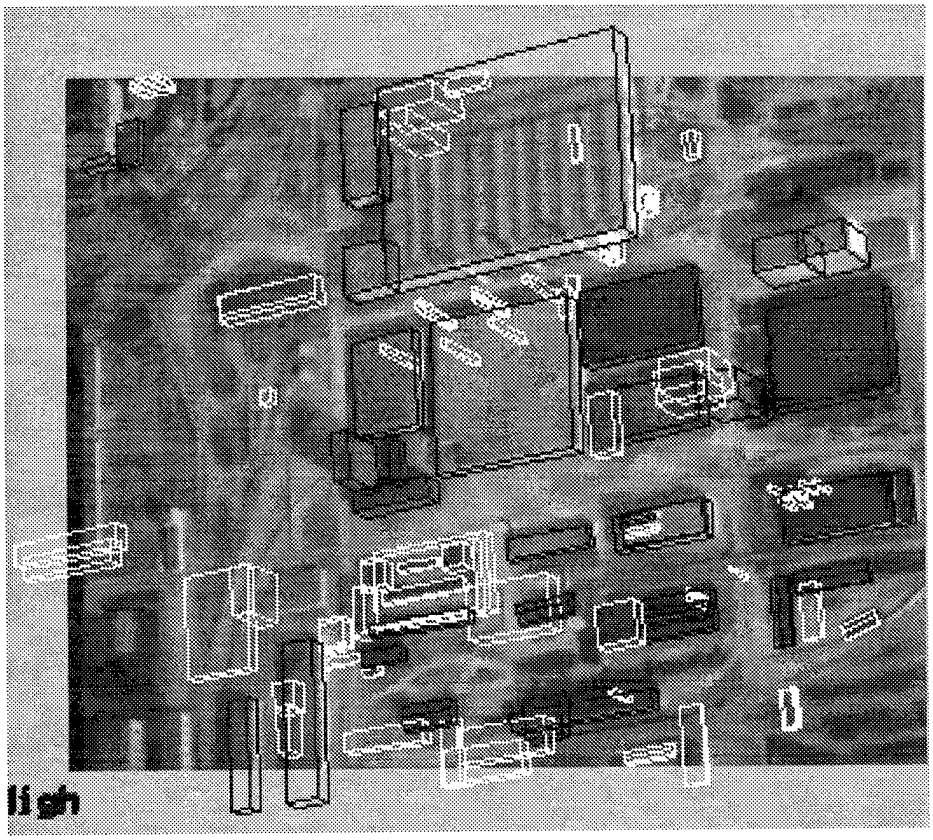


Figure 2.14 Missing buildings and false alarms (white outlines)

2.3.3.3 New Buildings

One important type of site change is the introduction of new structures. We have capabilities to construct models automatically, therefore we can suggest new additions to the site model. These techniques are applied to areas of interest using a single or multiple images, if available. The site model is used contextually to select the areas of search and to indicate existing modeled areas. The camera models and terrain models associated with the site and the various images are also used by these systems to derive viewpoint and illumination parameters automatically. An example of this task is shown later in Figure 2.17.

2.4 Discussion and Results

Next, we discuss the current capabilities of the system to direct attention to possibly changed structures. The processes that describe change in detail, and suggest modification (update) to the existing model objects are expected to require analysis using two or more images taken from different viewpoints and are currently under development.

2.4.1 Detection of Change

The ability to verify the presence of model objects (validation) in a new image is the base to determine whether changes have occurred. This process involves registering the model to a new image (or images), and establishing correspondences between the model and image elements at the object level. The quality and quantity of these correspondences give the first (or preliminary) indication of change to the structures. The indication of change currently comes in two forms:

- Confidence measures derived from match values reflect image support for a model object. Although low support may be caused by poor image quality and lack of contrast and occlusion, it can signify missing structures, substantially altered structures or incorrect modeling. These latter structures are presented in red color in the color figures below. Further use of context is expected to help interpret confidence values. Seasonal variations, for instance, may affect the appearance and visibility of building structures.
- Evidence of alteration. Model elements that correspond to image elements having greater extent represent preliminary indication of possible change. As previously discussed, some of these correspondences are explained as ambiguities caused by alignment of features. In the examples below, model buildings that exhibit extended correspondences are labeled with a circle placed on the center of mass of the structure. Note that this type of evidence, can be detected for any structure regardless of the confidence values assigned to it. The figures below show the evidence of change as thick cyan lines (white, in the mono-chrome versions of the results).

2.4.2 Choice of Parameters

Our system uses several parameters in its decision making processes at various levels. Choice of these parameters, of course, determines the quality of the results that are obtained even though we have attempted to make the system not very sensitive to them. Ideally, the parameter values should be based on a mathematical analysis of the algorithms and be a function of the parameters of the input images and any known parameters of the site. Unfortunately, such an analysis is difficult because of the complexities of the algorithms and estimating appropriate image parameters. In our system, we make decision parameters a function of the image parameters that are supplied with the images, such as the resolution. Additional refinements may be possible by using context of the site; our current system does not do so.

We have set the parameters by an informal analysis of the process and testing with a limited set of data. All of our examples use the same parameter values. We can not be sure that these parameters will also work well for other images and other sites. A more complete testing followed by a formal method of choosing optimal parameters, perhaps by a learning procedure, should help improve the choice. Our limited experience indicates that new images and new sites do not so much require changing of the parameters, but perhaps a need for additional reasoning steps to deal with situations not encountered in earlier tests.

2.4.3 Results

An example from the Fort Hood imagery supplied by the RADIUS program is shown below to demonstrate the current capabilities. The ability to detect change in the form of new structures (not in the model) is also demonstrated with an example below.

Figure 2.15 shows a portion of an image of Fort Hood, Texas, with the model overlaid. The image size is 7775x7720 pixels, and the 3-D site model contains 79 objects representing building structures. In this example, the image and the model are registered. The system therefore is applied to each model object separately using small image windows. Processing time is about 15 seconds per structure on a Sun sparc-10 workstation, running under the RCDE. We have grouped the previously discussed levels of confidence from 5 to 3 in the remaining results. The grouped levels are as shown below in Table 2.

Table 2. Grouped Confidence Levels

Match Value	**> 0.5**	**0.4 - 0.5**	**< 0.4**
Confidence	High	Medium	Low
Color	Green	Yellow	Red

The global results are described in Table 3. The table shows the number of building objects visible in the image and the distribution of validation confidence values. The color codes are in reference to the graphical results shown below in Figure 2.16. Note that 51 out of 79 objects in this image appear initially to have evidence of change.

Table 3. Summary of Results

Image	Visible Buildings	Validation Confidence			Possible Ambiguities		Non-changed Buildings			Changed Buildings			Missing Buildings		
		High (green)	Medium (yellow)	Low (red)	Detected	Resolved	Number of buildings	Reported non-changed	Reported changed	Number of buildings	Reported changed	Reported non-changed	Number of buildings	Reported missing	Validated
fhov927	79	54	13	12	51	33	54	53	1	14	13	1	11	11	0
Percent	100.0	68.3	16.4	15.1	100.0	64.7	100.0	98.1	1.9	100.0	92.8	7.3	100.0	100.0	0.0

The first step eliminates possible ambiguities. Thirty-three potential candidates for change are explained as ambiguities due to coincidental alignments and multiple matches. The remaining 18 are labeled "changed" based on evidence of extension of their physical dimensions. The "changed" building objects are shown with an orange circle in Figure 2.16. Table 3 shows that 14 of these are labeled correctly (Changed Buildings/Reported changed.) The remaining 4 are labeled correctly (Non-Changed Buildings/ Reported Changed) after the possibility of undermodeling or overmodeling is examined. In three of these cases undermodeling causes missing model elements on the roofs to be mistaken by detected image elements as extensions to portions of the structures. These ambiguities are handled correctly in our experiments. The one remaining case, reported incorrectly and shown in portion B of Figure 2.16, involves a coincidental alignment not handled by the current system.

One changed building, shown yellow in portion D of Figure 2.16, is not reported as changed (Changed Buildings/Reported non-changed). As the model is assumed to be in registration with the image, the system does not attempt to correct for translational errors beyond a few pixels. In this case, a slight error in positioning of the model prevented the appropriate correspondence with the image feature that would have otherwise signaled the possible extension.

Buildings that change considerably or are missing have poor image support, resulting in low validation confidence (the red buildings in the figures.) There are 12 of these, 11 of which were added by hand to test the "missing building" detection capability. The remaining one, shown in portion B of Figure 2.16, represents a significantly changed building. All these are labeled correctly as changed or missing.

Figure 2.17 shows the result of applying our monocular building detection system [10] to look for change in the form of new buildings (shown with bright outlines). The areas modeled are ignored. Typically the system would be instructed to locate new buildings in designated areas that are of interest, such as functional areas. The three buildings shown in white outlines are detected automatically.

2.5 Future Work

This system has been ported to SRI International for testing with operational imagery. A version that incorporates the RADIUS quick-look approach has been delivered to Lockheed-Martin for addition into the Radius Testbed System and to the National Exploitation Laboratory for evaluation and testing.

The current system operates in the 2-D domain of projected model structures onto the image viewpoint. Detailed description of change is expected to require the use of more than one image viewpoint, although the processing will continue to incorporate 2-D processing for simplicity and usefulness in regard to processing times required. We plan to explore the use of the verification mechanism by matching 3-D model features to 3-D features from multiple images or from a range sensor such as IFSAR. Our future work, however, will concentrate on giving detailed descriptions of detected changes to building structures, and to other structures of a permanent nature, such as roads and other transportation network objects.

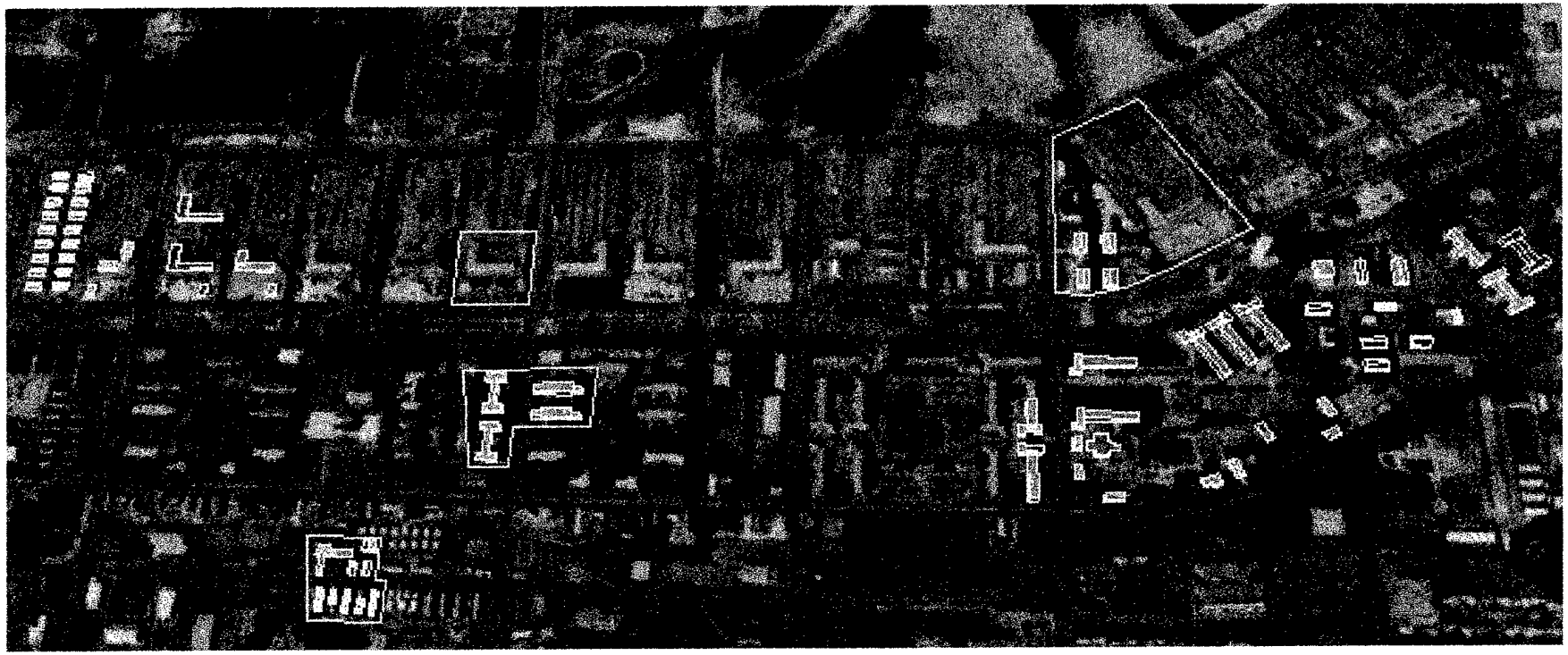


Figure 2.15 Portion of an image from Fort Hood, Texas

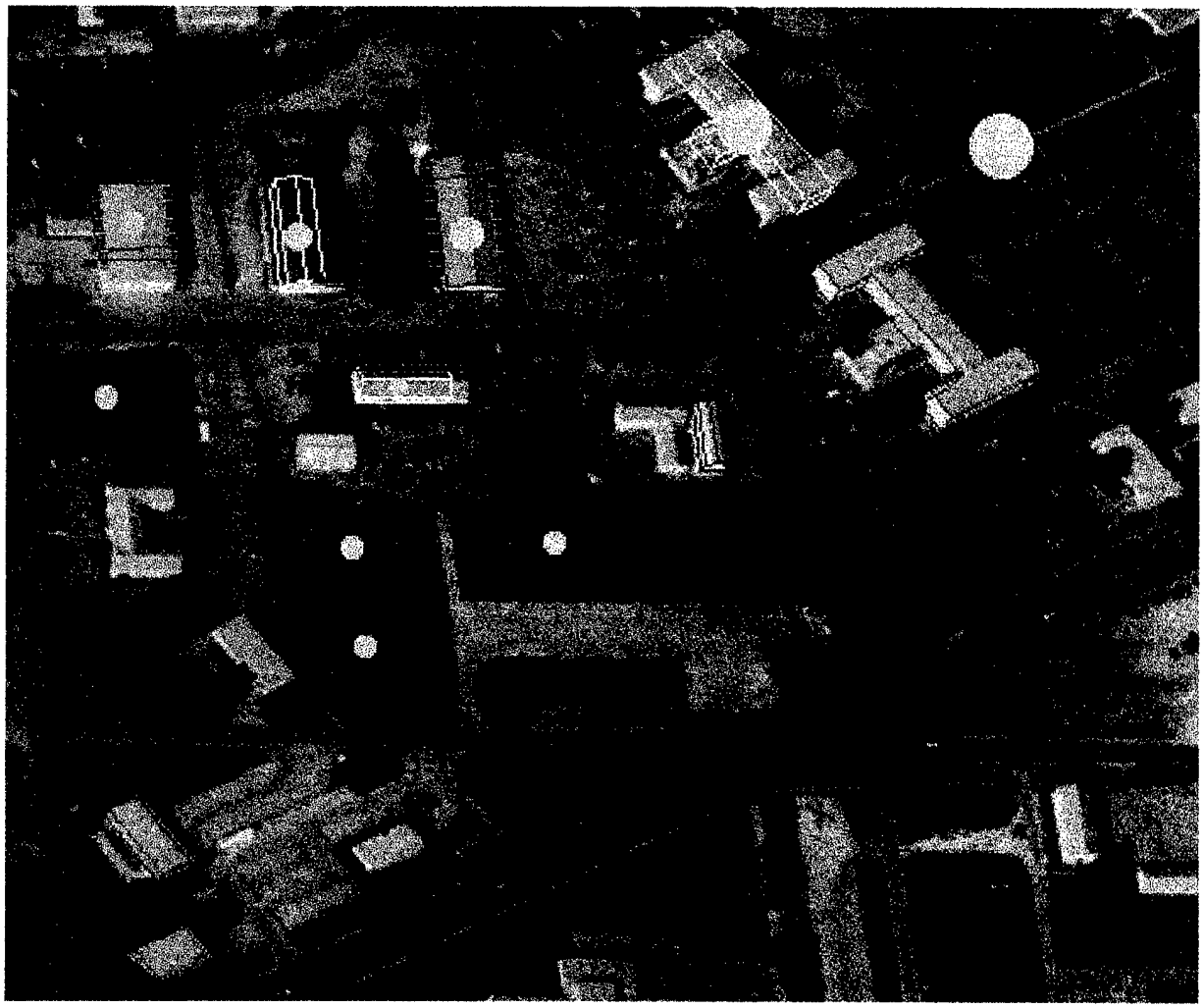


Figure 2.16 Portion A of the scene from Fort Hood

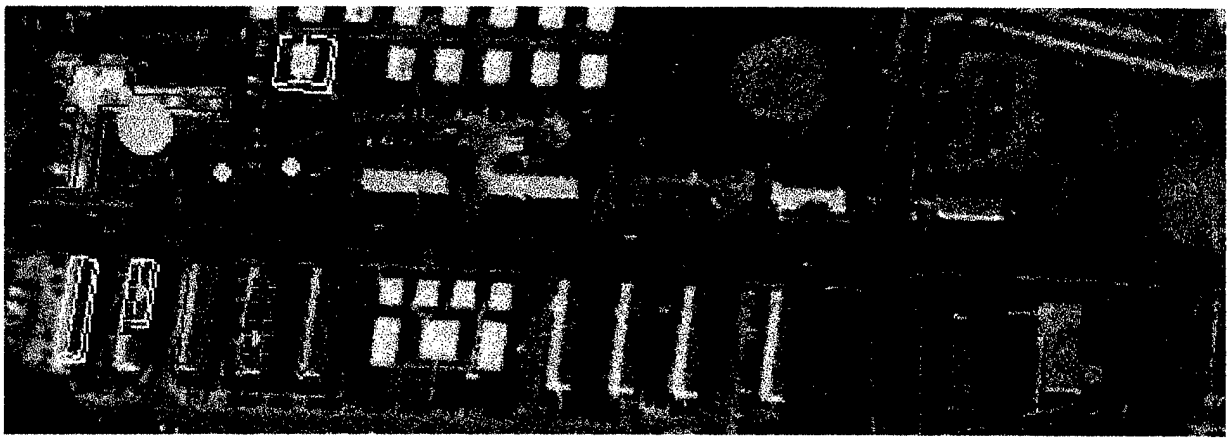


Figure 2.16 (continued) Another portion (B) of the scene from Fort Hood

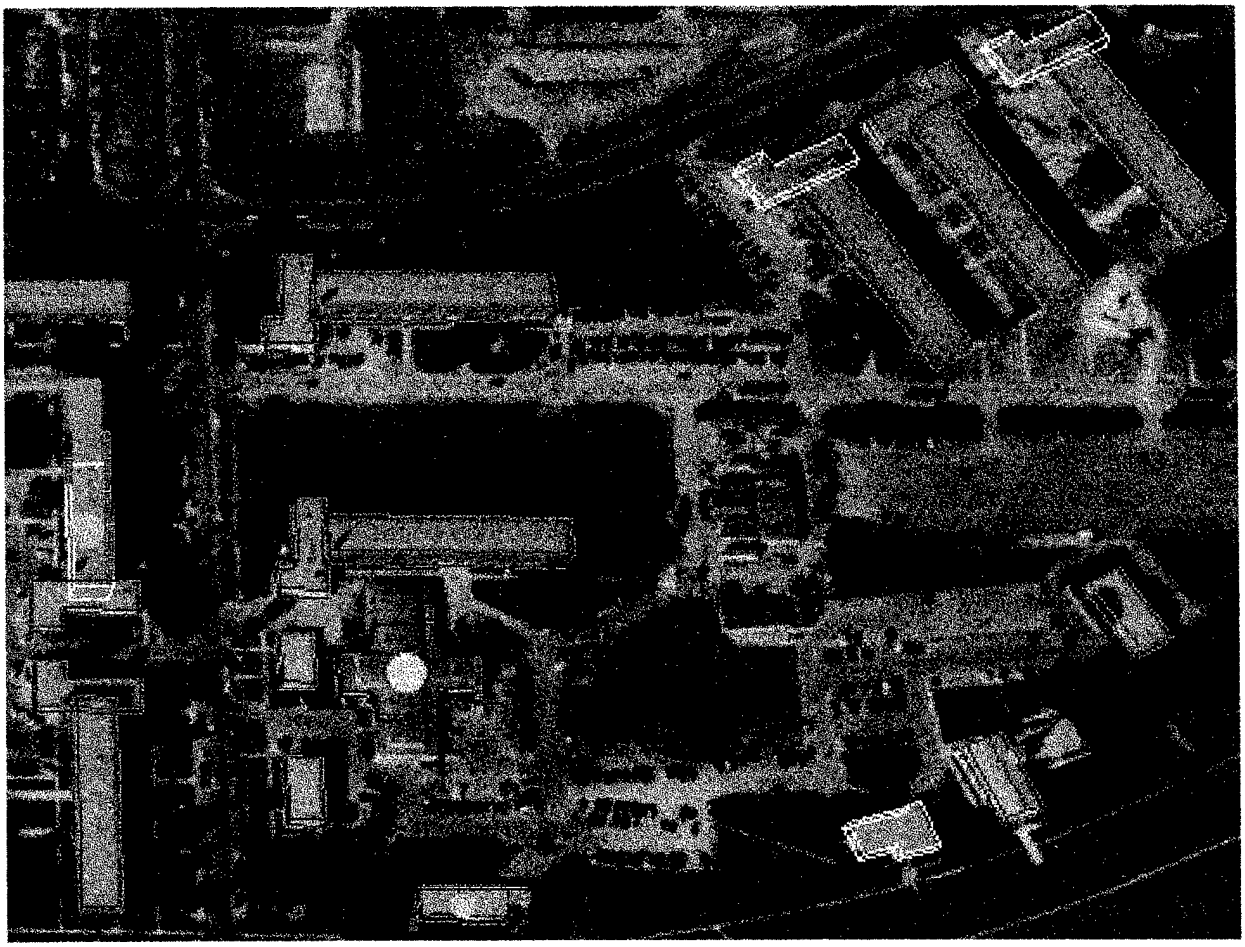


Figure 2.16 (continued) Another portion (C) of the scene from Fort Hood

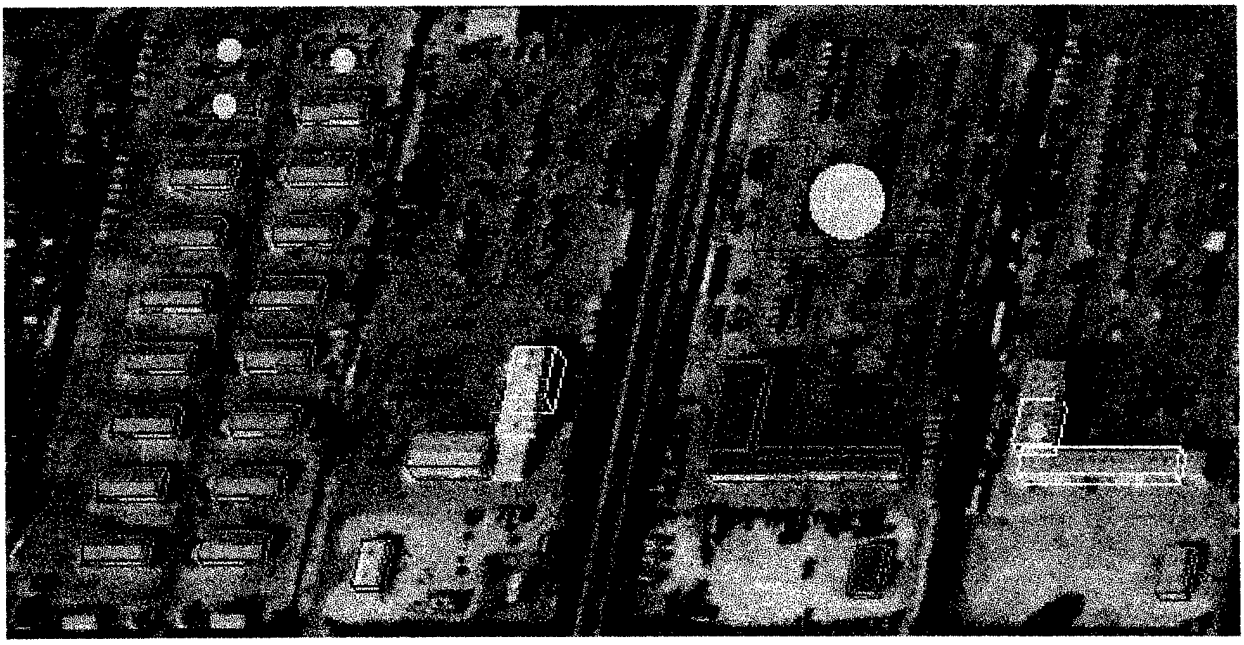


Figure 2.16 (continued) Another portion (D) of the scene from Fort Hood

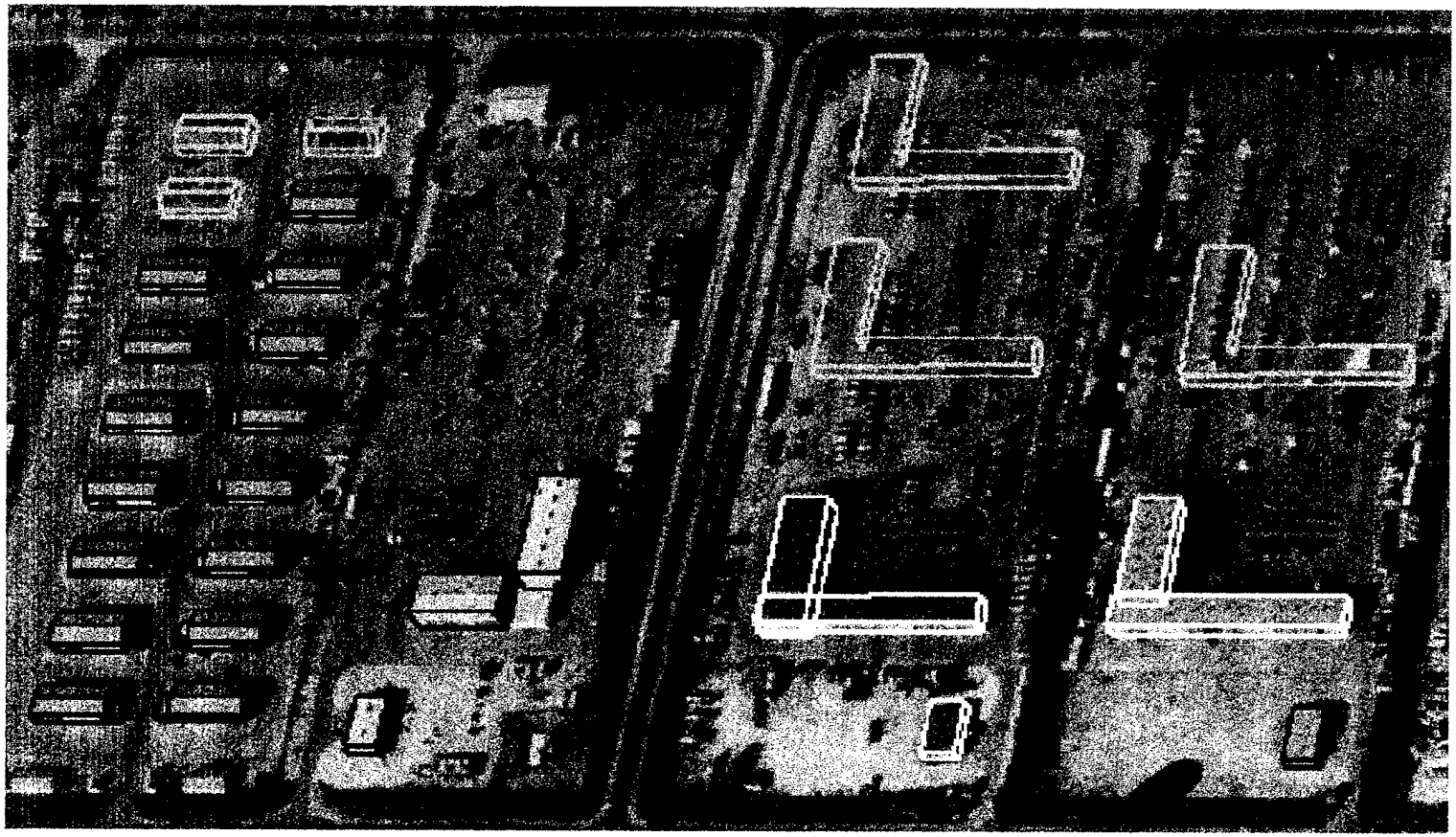


Figure 2.17 Change detection. New buildings are detected automatically

3 Automatic and Interactive Model Construction from a Single View

There are two major difficulties in inferring 3-D shape descriptions from a single intensity image. First, given an image, the system must find and separate objects from the background in presence of distraction caused by features from surface markings, vegetation, shadows, and highlights. This is the well-known "figure-ground" problem. The other difficulty is to construct 3-D from 2-D. Direct 3-D information is provided by a single intensity image, although the heights of the buildings can be estimated, under certain assumptions, from the shadow cast by them and by the visible walls.

Figure 3.1 shows the image of an L-shape building from Fort Hood. It is used as a running example in this report. Figure 3.1 also shows the line segments extracted from the image. The complexity of the task now becomes apparent. The building boundary is fragmented, and there are many extraneous boundaries from the surrounding objects. A building detection system must work under these conditions and also infer the 3-D shape.

The described system is designed to handle images from general viewpoints. It is easier to detect roofs of buildings from a nadir view image because of the shape constraint and better contrast. An oblique view image provides more 3-D cues than a nadir view, but many additional difficulties arise in the analysis process. First, the contrast between the roof and walls may be lower than the contrast between the roof and the ground, thus causing more fragmented boundaries. Second, small structures, such as windows and doors on walls, tend to interfere with the completeness of roof boundaries. Third, the projected shape of a building changes with the change of viewpoint. Fourth, the shadow of a building, which we use to verify the presence of a building and to estimate height, may be occluded by the building itself.

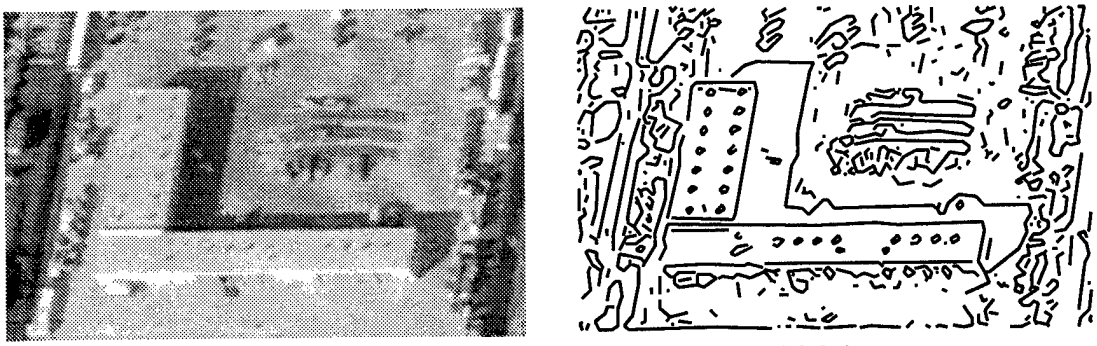


Figure 3.1 Image (left) and linear features (right)

There have been many methods proposed to solve the problem of building detection and description [9, 15, 19, 20, 21, 22, 23, 24, 25, 26]. The segmentation techniques usually rely on regions or edges extracted from the image. Region-based techniques construct closed curves that often do not correspond to the objects of interest. Simple edge-based techniques, such as contour tracing [20, 22, 25, 26], encounter

the problem of a rapidly growing search space. A more robust edge-based technique is the perceptual grouping technique [9, 15, 22]. For reconstruction of the 3-D information, most of the monocular systems use the corresponding shadow evidence of a building to infer the building height [9, 20, 21, 23].

Our system uses a perceptual grouping technique to generate roof hypotheses from the line segments detected from the image. A feature hierarchy is created by grouping line segments into parallel lines, u-contours, and parallelograms, which correspond to roof hypotheses. The skewness of roof hypotheses is handled according to the viewpoints. A selection process selects good hypotheses for verification based on the local evidence and the global relationships. Some 3-D cues, such as OTVs (Orthogonal Trihedral Vertex) and matched shadow corners, are used in the selection process. Shadow and wall evidence is used to verify the selected hypotheses. The use of both shadow and wall evidence makes the verification process generate more reliable results and makes the system more robust. The 3-D information, that is, the building height, is inferred from the shadow and wall evidence. Figure 3.2 shows the block diagram of the system.

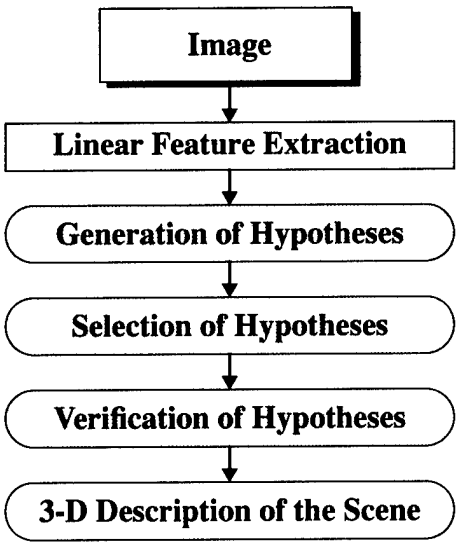


Figure 3.2 Block diagram of the automatic system

Our system design philosophy has been to make only those decisions that can be made confidently at each level. Thus, the hypothesis generation process creates as many hypotheses as feasible. The hypothesis selection process favors keeping hypotheses that may be viable. The hypothesis verification process has the most global information and therefore can make more informed decisions.

The automatic system often works quite reliably under certain conditions, but the results are not perfect and some manual editing is still required. A semi-automatic (interactive) system that requires minimal input from a user to perform this task has been developed on top of the automatic system to make use of the results and functions of the automatic system.

A variety of interactive systems has been built for site model construction [6, 27]. The amount of interaction varies from almost complete manual operation with an operator locating all the significant features, to one in which the operator selects a parametric model or a rough outline which is then fitted to image data under operator control. In all such cases, the task of the machine is limited to that of bookkeeping, simple geometric calculations, or some form of error minimization. No perceptual capability of the ma-

chine is utilized and the operator is required to provide a large number of inputs. We will refer to such systems as *computer-assisted manual systems.*

We have proposed an alternative strategy to combine the activities of the operator and the machine by taking advantage of the perceptual abilities of the machine.

The automatic system usually fails to detect a building because the evidence is weak or there is an incorrect determination of one or two sides of the roof hypothesis. In some cases, very simple guidance from the user, merely indicating that in fact a building is present in the vicinity, suffices for the semi-automatic system to find one on its own. The user can use the interactive system to correct the wrong side(s) of a hypothesis easily and then the system can verify the hypothesis again and find the correct height automatically. The goal is to provide a minimum amount of user input to the machine and allow the machine make the decisions. The methodology does allow for more detailed interaction with the system, in stages, and as necessary. In the worst case, the system requires the user to provide all the information, as in most manual systems. However, this capability is seldom needed in this system.

This system makes the following assumptions: the projection is locally-weak perspective, viewing angles and sun angles are known, roofs are flat and rectilinear, walls are vertical, and shadows fall on flat ground.

The system has been tested on several examples of the modelboard images and Fort Hood images provided by the RADIUS program. Some results and an evaluation of the results are presented later in this report.

3.1 Generation of Hypotheses

First of all, the system uses an edge detector to extract intensity linear features from the image. A perceptual grouping process is then used to generate roof hypotheses by constructing a feature hierarchy from the linear features. The feature hierarchy, which includes linear, parallel, U-contour (portions of parallelogram) and parallelogram features, encodes the structural relationships specific to the projection of rectangular shapes, presumably corresponding to the visible flat roof surfaces. A perceptual grouping process (see Figure 3.3) is used to group low-level features into high-level features to form the feature hierarchy where linear features are grouped into parallel features, linear features and parallel features are grouped into U-contour features, and U-contour features are grouped into parallelogram features which are the roof hypotheses.

A group of close parallel segments represents a linear structure at a higher granularity level than the segments. The segments-folding process groups close parallel segments into a sub-segment. The generated sub-segment has a length and an orientation derived from the contributing segments. The system then detects L-junctions and T-junctions among these sub-segments. The junctions are used to break sub-segments into edges. A colinearization process groups colinearized edges into a longer edge to overcome the problem of fragmented line segments generated by the low-level vision process.

Man-made structures in urban scenes, such as buildings, roads, and parking lots, are often rectilinear. Therefore, parallel sides are present in these structures. Furthermore, the projection of 90 degree corners of these structures can be computed as a function of the orientation of one side of the corner (α), swing angle (θ) and tilt angle (γ). Swing angle and tilt angle can be derived from the camera model of the image. Figure 3.4 and equation (3.1) show the angle constraint of the projection of a 90 degree corner of a flat roof.

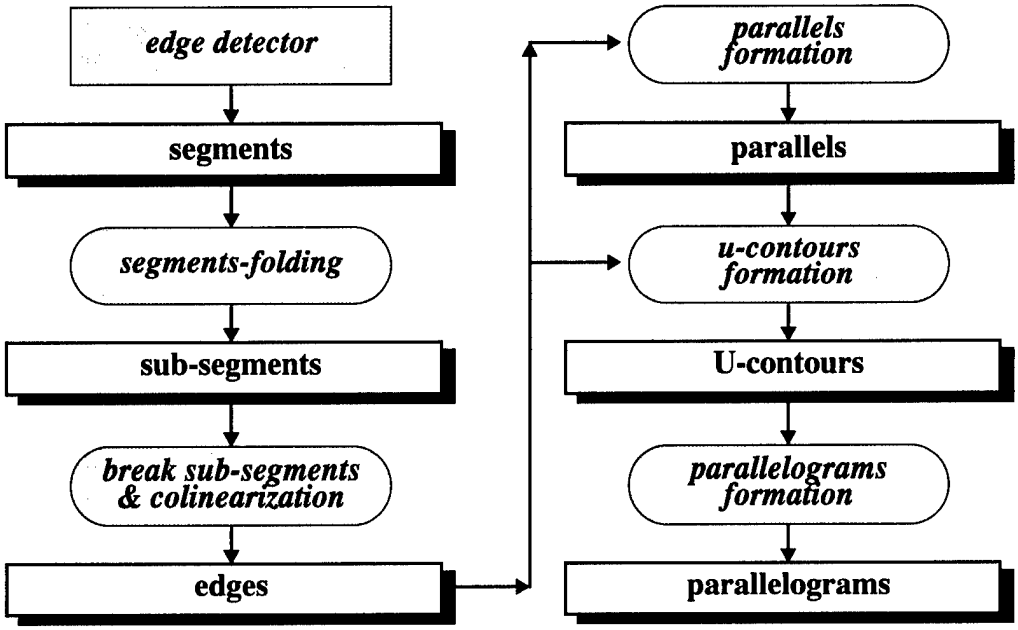


Figure 3.3 Hierarchical perceptual grouping

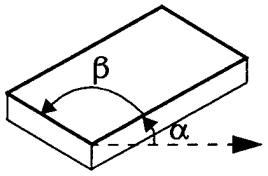


Figure 3.4 Right angle constraint

$$\beta = \operatorname{atan}(\mu, \nu)$$

$$\text{where} \begin{cases} \mu = \cos^2(\alpha+\theta)\cos(\gamma) + \dfrac{\sin^2(\alpha+\theta)}{\cos(\gamma)} \\ \nu = \sin(\alpha+\theta)\cos(\alpha+\theta)\left(\cos(\gamma) - \dfrac{1}{\cos(\gamma)}\right) \end{cases} \tag{3.1}$$

As a consequence, the formation of parallel features is an important step in making roof hypotheses. The system uses two edges, which are parallel and *aligned* to each other, as a trigger for making a parallel feature. Two parallel lines are *aligned* if the angle between the direction of the parallel lines (α) and the direction of the line connecting the end points of the two lines satisfies the right angle constraint, that is, equation (3.1). The alignment of the two lines in a parallel feature also strongly suggests the presence of a line along the aligned ends of the parallel feature. Therefore, the line (base line of the U-contour) and the

parallel feature are grouped into a U-contour feature. Even if the base line of the U-contour does not have any edge evidence, we still hypothesize it and generate the U-contour. Two U-contour features that make a closure are grouped into a parallelogram. The parallel features of these two U-contour features must be parallel and close to each other. As a result of the right angle constraint, the base lines of these two U-contour features must be parallel. The parallelogram features correspond to roof hypotheses and the following two processes select and verify these roof hypotheses.

Figure 3.5 shows all of the 74 roof hypotheses generated by the system from the 1,049 linear features in Figure 3.1.

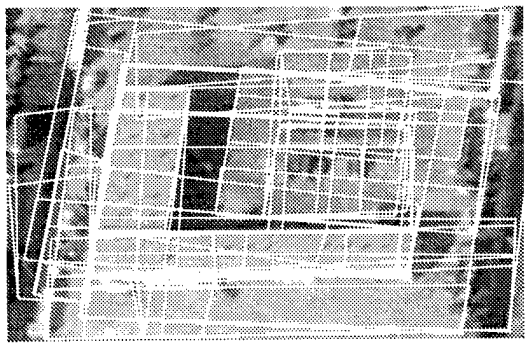


Figure 3.5 All hypotheses

3.2 Selection of Hypotheses

After the formation of all reasonable roof hypotheses, a selection process is applied to choose hypotheses having strong evidence of support, with minimum conflict among hypotheses. Based on the local and global supporting evidence of hypotheses, a rule-based selection process selects promising hypotheses for verification. This process greatly decreases the number of hypotheses to be verified, and therefore reduces the run time of the time-consuming verification process.

The selection process uses two kinds of criteria: **local selection criteria** and **global selection criteria** (see Figure 3.6). Local selection criteria determines whether or not a parallelogram is *good* based on the local supporting evidence, such as lines, corners, and their spatial relations. A score for each parallelogram hypotheses is computed by using all local selection criteria. Only *good* parallelograms, that is, the parallelograms whose scores are greater than a given threshold value, are retained for global selection. It is possible that some of the good parallelograms retained after the local selection are almost the same or overlapped with each other. Global selection criteria selects the best consistent parallelograms from the *good* parallelograms. More than one global selection criteria can be used to filter out some hypotheses according to each global selection criterion.

The local selection criteria are derived from both ***positive evidence*** and ***negative evidence*** of existence of a roof. The positive evidence includes the presence of edges, corners, parallels, OTVs and matched shadow corners (see Figure 3.7). The negative evidence includes the presence of lines crossing any side of a parallelogram, existence of L-junctions or T-junctions in any side of a parallelogram, existence of overlapped gaps on opposite sides of a parallelogram, and displacement between a side of a parallelogram and its corresponding edge support (see Figure 3.7).

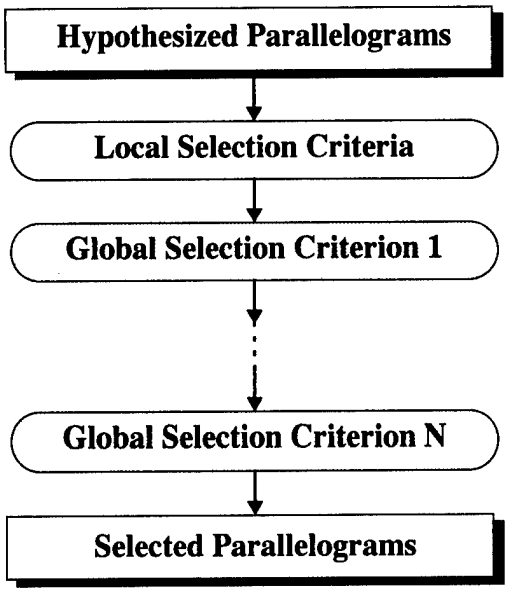


Figure 3.6 Selection process

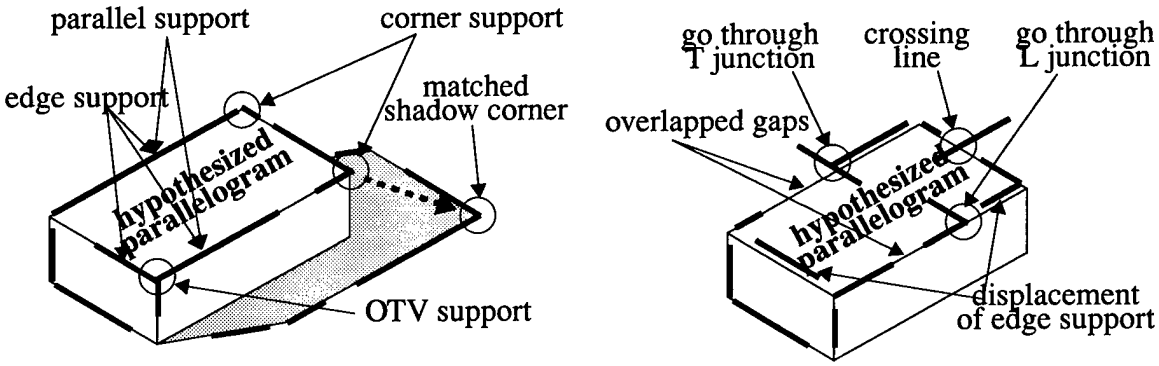


Figure 3.7 Evidence. Positive (left) and negative (right)

Negative evidence is as important as positive evidence, because it helps in the removal of those parallelograms which are less likely to correspond to building roofs. Each kind of evidence is formulated into an evaluation criterion. A weight is assigned to each evaluation criterion according to the probability of existence of a building under the presence of the evidence, which the criterion evaluates. We can emphasize the importance of an evaluation criterion by assigning a higher weight to it. For each parallelogram, the score evaluated by the local selection criteria is the weighted sum of all the local evaluation criteria. If the score is greater than a given threshold value, the parallelogram is considered as a *good* hypothesis and retained for global selection.

Good parallelograms surviving local selection may compete with each other. For example, some parallelograms could share the same edges or corners support and some parallelograms might overlap with each other. The goal of global selection is to select a minimum set of parallelograms which best describe the rectangular composition of the scene. Global selection criteria examine overlapping parallelograms and choose one if appropriate. The selection is based on relative properties of each parallelogram, the amount

and kind of overlap, and whether they share support or not. If a parallelogram does not overlap with any other parallelogram then it is not in competition, and it is retained. There are three global selection criteria in this system: elimination of duplicated hypotheses, selection in containment situation, and selection in overlap situation.

The system selects 3 hypotheses out of the 74 hypotheses. The selected hypotheses is shown in Figure 3.8.

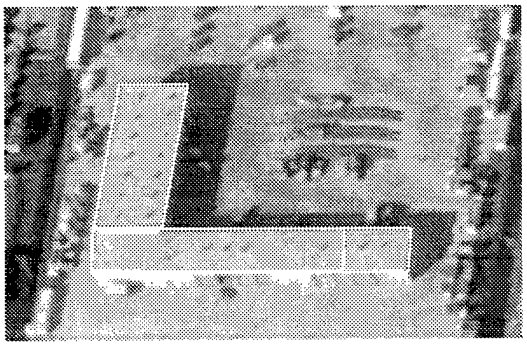


Figure 3.8 Selected hypotheses

3.3 Verification of Hypotheses and Inference of 3-D Shape

The purpose of verification is to confirm that the selected hypotheses correspond to buildings. The existence of wall or shadow evidence increases our confidence that the hypothesis is actually a part of a 3-D structure. Also, such evidence provides this system with the 3-D information required to create the 3-D model of the structure.

Figure 3.9 shows the projection of a typical building and illustrates some of the parameters used by our system. Given a roof hypothesis, we do not know if the hypothesis actually corresponds to a building roof. Even if it did, the height of the building is unknown. However, we know that the building height, *H*, is within a certain range.

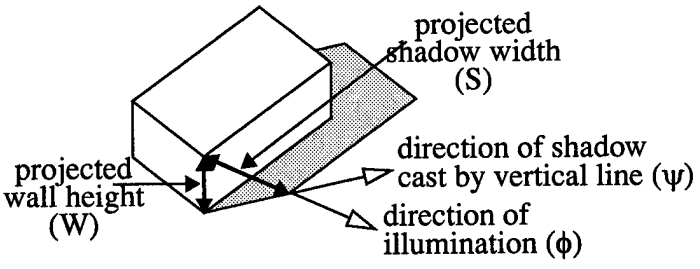


Figure 3.9 Wall height and shadow width

Assume that the image resolution is R (pixels/meter). The projected wall height, W, can be computed from the building height and the viewing angles (the swing angle, θ, and the tilt angle, γ) by equation (3.2).

$$W = H \cdot R \cdot \sin\gamma \tag{3.2}$$

Also, the projected shadow width, S, can be computed from the building height, the viewing angles, and the sun angles (the direction of illumination, ϕ, the direction of shadow cast by a vertical line, ψ, and the sun incidence angle, i) by equation (3.3).

$$S = \begin{cases} H \cdot R \cdot \tan i & \text{when } \gamma = 0 \\ \dfrac{H \cdot R \cdot \sin(i+\gamma)}{\cos i} & \text{when} \begin{cases} \gamma \neq 0 \\ \psi = \phi = 270° - \theta \end{cases} \\ \dfrac{H \cdot R \cdot \sin(i-\gamma)}{\cos i} & \text{when} \begin{cases} \gamma \neq 0 \\ \psi = \phi = 90° - \theta \end{cases} \\ \dfrac{H \cdot R \cdot \sin(\gamma - i)}{\cos i} & \text{when} \begin{cases} \gamma \neq 0 \\ \psi = \phi + 180° \end{cases} \\ \dfrac{H \cdot R \cdot \sin\gamma \cdot |\cos(\psi+\theta)|}{|\sin(\psi-\phi)|} & \text{otherwise} \end{cases} \tag{3.3}$$

Therefore, we can search for wall and shadow evidence, such as lines and corners, in a certain neighborhood of a given roof hypothesis. Each piece of evidence contributes to the confidence of a hypothesis as explained below (in Sections 3.3.1 through 3.3.3). Hypotheses with high confidence are considered to be verified.

A containment and overlap analysis process is then applied to resolve some of the remaining ambiguities, such as when one verified hypothesis contains or overlaps with another verified hypothesis. This process is explained in Section 3.3.4. A block diagram of the verification process is shown in Figure 3.10.

3.3.1 Wall Verification Process

Generally, some walls of buildings should be visible in an oblique view. As obliqueness increases wall information becomes more useful and shadow information becomes more difficult to handle, if it is available at all. We assume that walls are vertical.

The purpose of the wall process is to find wall evidence at every possible building height for each roof hypothesis. Given the viewing angles and a possible building height, we can estimate wall boundary for a roof hypothesis. All evidence around the wall boundary is collected and a score is computed for the wall evidence.

With the knowledge of the minimum and maximum heights of buildings, the search for wall evidence is limited to a certain range. The system can either do an exhaustive search over the range or do a smart search. The smart search is performed by taking samples within the search range to locate some evidence

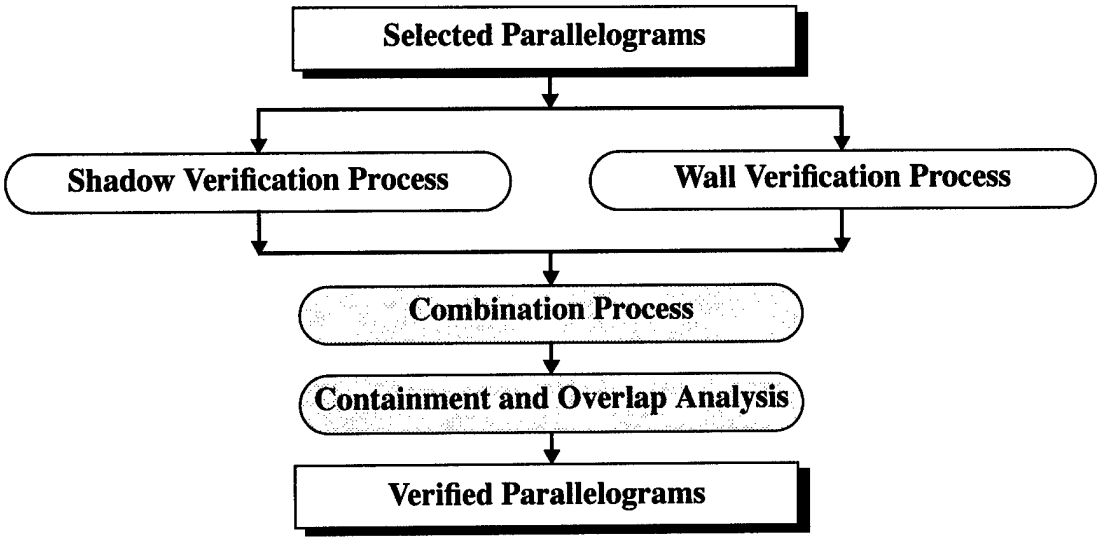


Figure 3.10 Verification process

first and then doing a search only on those positions where the chance of finding wall evidence is high. The smart search does not always find the best solution; however, with appropriate sampling, it could be fast and find the best solution most of the time. Currently we are using an exhaustive search algorithm; the smart search algorithm will be implemented in the near future.

Given a roof hypothesis, view angle information allows the determination of the visible sides of the building. The swing angle gives the vertical direction from which building sides are hypothesized. The projected wall height (see Figure 3.9) can be computed by equation (3.2), given a possible building height, H. We delineate the wall boundary and activate a search process to collect all evidence along the delineated wall boundary. For each possible building height, a set of corresponding wall evidence is collected for evaluation. Figure 3.11 shows the search of wall evidence at several possible building heights.

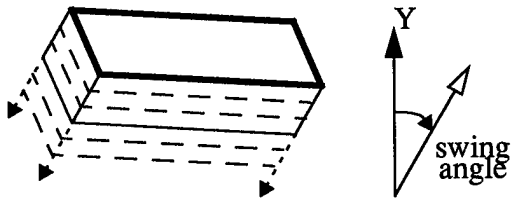


Figure 3.11 Search for wall evidence

The evaluation process evaluates the wall evidence collected from the previous step. Basically the score is a weighted sum of the evidence of ground-boundary, vertical-boundary, and corners. Equation (3.4) is the evaluation function of the wall evidence of a hypothesis p at the building height H. k_i is an evaluation function for the i-th component of the wall evidence and v_i is the corresponding weight.

$$W(p,H) = \sum_i v_i \cdot k_i(p,H) \tag{3.4}$$

3.3.2 Shadow Verification Process

The use of shadow evidence to verify hypotheses is more complicated in oblique views than in nadir views, for the shadow may be occluded by the building itself (see Figure 3.9).

The shadow verification process tries to establish correspondences between shadow casting elements and shadows cast. We assume that shadows fall on flat ground. The shadow casting elements are given by the sides and junctions of the selected roof hypotheses. The shadow boundaries are searched for among the lines and junctions extracted from the image.

There are a number of difficulties that prevent the accurate establishment of correspondences. Building sides are usually surrounded by a variety of objects, such as loading ramps and docks, grass areas and sidewalks, trees, plants and shrubs, vehicles, and light and dark areas of various materials. Occlusion of the shadow by the building itself or by nearby buildings may make the shadow region irregular and make the shadow evidence difficult to extract. To deal with these problems we have adopted some geometric and projective constraints and special shadow features.

The potential shadow evidence is extracted from the linear features of the image and the knowledge of the sun angles: lines parallel to the projected sun rays in the image may represent potential shadow lines cast by vertical edges of 3-D structures; lines having their dark side on the side of the illumination source are potential shadow lines. Junctions among the potential shadow lines are potential shadow junctions, and neighborhood pixel statistics give relative brightness.

Given the sun angles and viewpoint angles, we know which sides of a roof will cast shadow and which part of the shadow will be occluded by the building itself. The shadow is cast along the direction of illumination. The projected shadow width (see Figure 3.9) can be computed by equation (3.3) given a possible building height, H. We can then delineate the projected shadow region in 2-D with the appropriate removal of the self-occluded shadow region. The shadow verification process collects all potential shadow evidence along the delineated shadow boundary. For each possible building height, a set of corresponding shadow evidence is collected for evaluation. Figure 3.12 shows that the system searches for shadow evidence at several possible building heights.

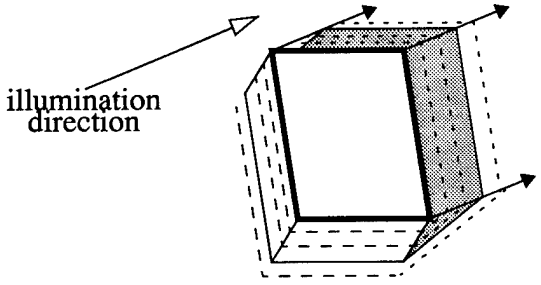


Figure 3.12 Search for shadow evidence

The shadow evidence associated with each possible building height is evaluated and given a score as a weighted sum of the evidence of shadow lines cast by roof, shadow lines cast by vertical lines, shadow junctions and the shadow region statistics. Equation (3.5) is used to compute a score for the shadow evidence of a hypothesis p at the building height H, where h_i is an evaluation function for the i-th shadow evidence and u_i is the corresponding weight.

$$S(p,H) = \sum_i u_i \cdot h_i(p,H) \tag{3.5}$$

3.3.3 Combination of Shadow and Wall Evidence

For each hypothesis, p, the previous two steps calculate a shadow score, $S(p, H)$, and a wall score, $W(p, H)$, for the building height, H. Next these scores are combined by equation (3.6).

$$\begin{aligned} &C(p,H) = S(p,H) + W(p,H) - S(p,H) \times W(p,H) \\ &\text{note that } \; 0 \le S(p,H), W(p,H) \le 1 \end{aligned} \tag{3.6}$$

For each hypothesis, p, the building height that gives the highest combined score is considered to be the estimated building height of the hypothesis and the corresponding score is called the confidence value of the hypothesis. Equation (3.7) shows the definitions of estimated building height and confidence value.

$$\begin{aligned} &C_p = C(p, H_p) = \text{Max } C(p,H) \\ &\text{where} \begin{cases} H_p : \text{estimated building height for hypothesis } p \\ C_p : \text{confidence value of hypothesis } p \end{cases} \end{aligned} \tag{3.7}$$

If the confidence value of a hypothesis is greater than a given threshold value, the hypothesis is considered verified. The use of certainty theory in equation (3.6) allows our system to verify a hypothesis based solely on the wall evidence or shadow evidence. This makes it possible to handle the cases of imperfect wall or shadow evidence.

3.3.4 Containment and Overlap Analysis

The wall and shadow verification processes examine each hypothesis individually and do not analyze any relationship among them. Thus, some verified hypotheses might contain others or they may overlap with each other. A containment and overlap analysis of the verified hypotheses is used to resolve the problem of having more than one building in the same 3-D space.

For example, in Figure 3.13, hypothesis (EFCD) is contained by hypothesis (ABCD). They both have enough wall and shadow evidence to allow them to be verified. It is not necessary that the outer hypothesis is always the preferred one; sometimes, it may contain elements other than those belonging to a building. The system makes a choice by examining the evidence of the roof, the wall, and the shadow boundaries on the *non-shared* sides of the conflicting hypotheses. In the example shown in Figure 3.13, the outer hypothesis has more supporting evidence on the non-shared side (AB), so the inner hypothesis is discarded.

After the containment analysis, the system applies overlap analysis to the retained hypotheses. The idea is the same as the containment analysis. The system examines the *non-shared* parts of the hypotheses and decides whether the evidence is strong enough to keep the hypothesis. Currently the system uses a very simple algorithm in the overlap analysis. If two overlapped hypotheses have the same height and have a large overlapping area, the system removes the one with lower confidence value, otherwise both hypotheses

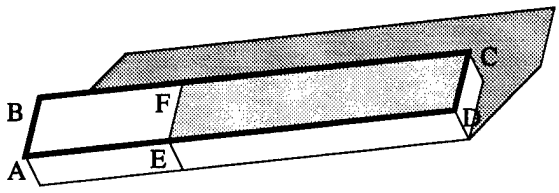


Figure 3.13 Containment analysis

are retained. A more sophisticated algorithm is required to analyze the case where more than two hypotheses are involved. The decision may have to choose the best combination of hypotheses among several groups of hypotheses.

Figure 3.14 shows the two hypotheses verified by the system from the three selected hypotheses in Figure 3.8. The upper part of the structure is verified because it has a clear shadow boundary, although no wall evidence can be found. The lower part of the structure has fragmented wall boundaries and imperfect shadow boundaries, but the system is able to spot the small pieces of evidence and verify it. Figure 3.14 shows the verified hypotheses in 3-D wire frame format.

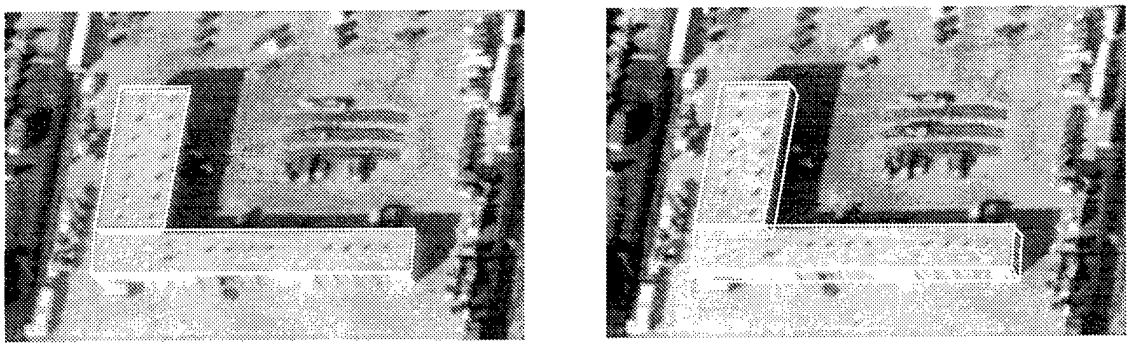


Figure 3.14 Verified roof hypotheses (left) and 3-D wire frame model (right)

3.3.5 3-D Description of Buildings

The 3-D information of the verified buildings, that is, the roof hypothesis and the estimated building height, together with the camera model and the terrain model of the scene are used to generate a 3-D wire frame model (see Figure 3.14) of the scene. The textures inside the roofs and visible walls of verified buildings are painted onto the corresponding surfaces in the 3-D wire frame model. The textures of the ground surface in the input image are painted onto the ground surface of the 3-D wire frame model also. This 3-D wire frame model can be viewed from an arbitrary viewpoint. The transformation that projects the 3-D scene onto a 2-D screen for viewing can then be used to collect the pixel values from the 3-D wire frame model and use them to render the projected image (see Figure 3.15).

3.4 The Interactive System

The performance of the automatic system on several examples is presented in section 3.5. While this system performs well under many conditions, there are several situations that cause it to fail to find a building or a correct description of the building. An interactive system has been developed to correct these errors.

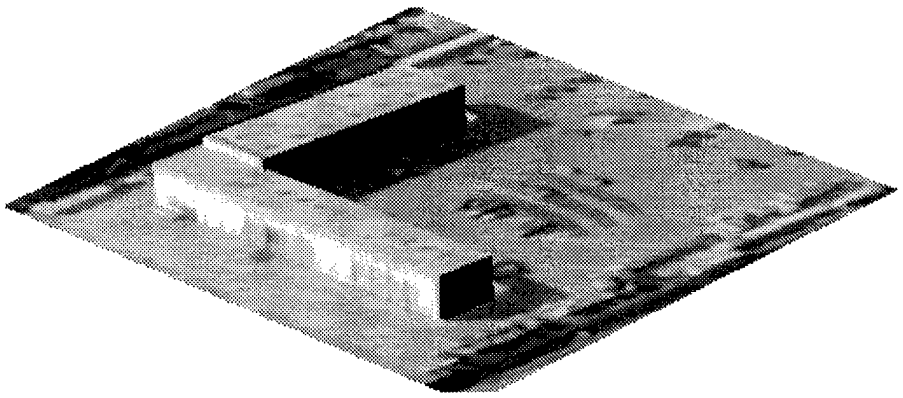


Figure 3.15 Rendered image from another viewpoint

The goal of the interactive system is to minimize the required interaction by using the partial results of the automatic system where possible. The process of interaction can be divided into two parts, *initial interaction* and *corrective interaction* (see Figure 3.16).

- Initial Interaction (qualitative)
  First the user classifies the detection problem, such as dark areas, poor contrast, occluded buildings, occluded shadows, or partly detected L- or T-buildings. The classification scheme depends on the performance of the automatic system. This qualitative information is useful to constrain the search for new hypotheses. Although the classification is, in general, not unique (several problems may occur at the same time), it is unlikely that a correct hypotheses will be rejected as long as the classification is correct. The second qualitative step is a rough localization of the missing building. This can be, for example, any point on the roof (it is possible to automate this step by clustering rejected hypotheses). After the initial interaction, the most likely hypothesis can be established by the additional information provided by the user.
- Corrective Interaction (quantitative)
  If the hypothesis established in the first step is (partly) wrong, the user must correct the sides or corners of the building model. For example, if one roof-side is incorrect, the user can either drag the line to the desired location or select a line in the image that refers to the roof-side. After each single correction, both the verification and fitting of the parallelogram and the determination of the building height will be carried out automatically. This step can to be repeated until the operator is satisfied with the result. In the worst case, the complexity of interaction here is the same as in a manual system.

To use intermediate results and the functions of the automatic system, we have to find a stage where the user inputs can be readily used. By analyzing the automatic system, the earliest convenient intermediate results are all the roof hypotheses generated by the automatic system. At this stage, the information, which is computed by the feature extraction and perceptual grouping, is still available and a unique representation for a possible building can already be obtained.

3.4.1 Initial (Qualitative) Interaction

The input for this step is the qualitative information (indication of the problem and a rough location of a missing building) from the user and all the roof hypotheses generated by perceptual organization. Ac-

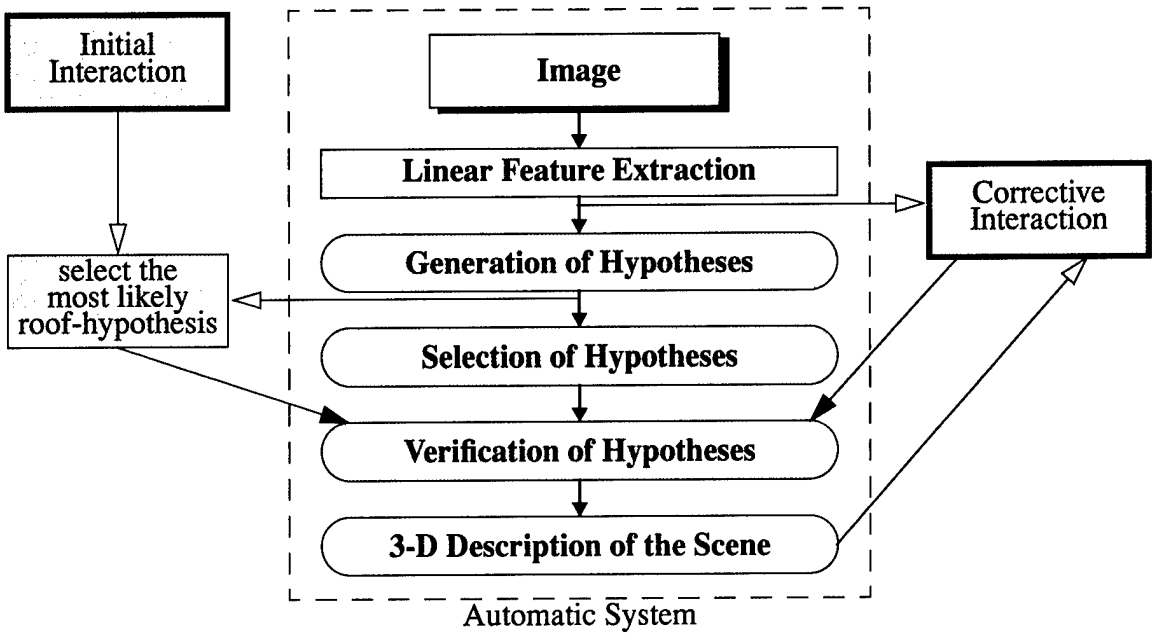


Figure 3.16 Interaction system embedded in automatic system

cording to the specified location, a local subset of hypothesized parallelograms is established, from which the most likely hypothesis according to the detection problem is selected. When no detection problem is specified, therefore no specific knowledge of the scene is known, the system uses the one with the highest hypothesis score, which is computed by the selection process of the automatic system.

A set of parallelogram-patterns is associated with each kind of detection problem, which classifies the parallelogram hypotheses that can occur for the problem. An example of a pattern is a parallelogram, in which one roof-side is wrong by a translation (because there were no edges detected at this roof side), and all other sides and the angles are correct (see Figure 3.17). Another example of a pattern is a complete match between the parallelogram hypothesis and roof, which would lead to a correct guess after the initial interaction.

This set of patterns must be established by the designer of the system after an analysis of system failures. Once a class of problems is selected, the probability of being the missed hypothesis is assigned to each parallelogram hypothesis according to the set of patterns: observation x_i for each pattern j is collected and transformed to a number ω_i which can be related to the associated likelihood. The observation can be represented either as a real number, an integer, or a boolean.

$$\omega_i = \frac{1}{2}\left(\frac{x_{ij} - \bar{x}_{ij}}{\sigma_{ij}}\right)^2 \quad \text{for real numbers}$$
$$\omega_i = -\ln P(x_{ij}) \quad \text{for integers/boolean} \tag{3.8}$$

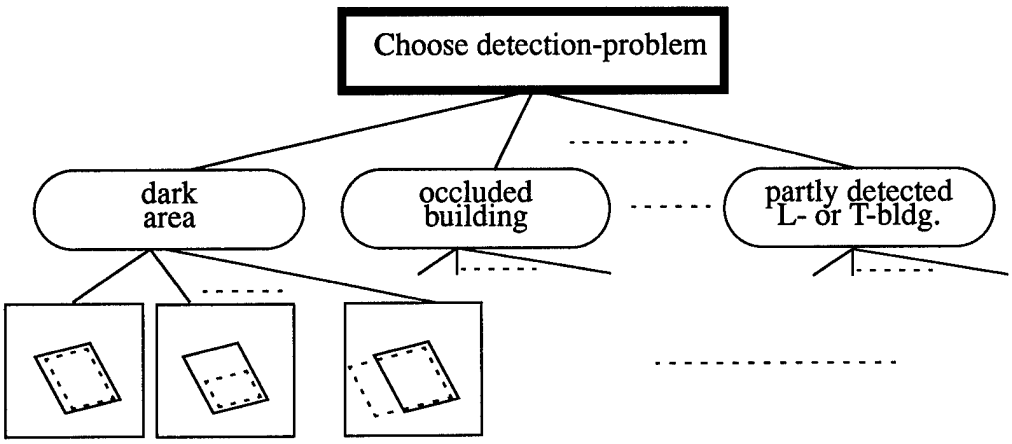


Figure 3.17 Example for classes of detection problems and their patterns

These formulas are derived by assuming Gaussian distribution. The parameters, x_{ij}, σ_{ij}, and $P(x_{ij})$ (mean value, standard deviation, and probability of observation x_{ij}), must be determined either theoretically or empirically.

For each pattern $e^{-\Sigma\omega_i}$ is proportional to the likelihood, so that the most likely pattern for each parallelogram can be chosen. Similarly, the most likely hypothesis for the roof of the missing building is selected by comparing the ω of the most likely pattern of each parallelogram.

An advantage of this selection method is that the system can, because of the selected pattern, give a prediction with a certain probability as to whether a corrective interaction is necessary, and where the interactions must be made.

Example: dark buildings

As an example, we want to introduce possible evidence for the problem-class of dark buildings. Usually the boundary between the shadow and the roof is difficult to detect. The image edges of two sides of the roof are at best only partly visible. Three observations are sufficient to select the best hypothesis available after perceptual organization: evaluation of the parallelogram-corners, of gray value changes at the roof boundaries, and of the overall average gray-level. Two patterns are used, one where all sides are correct and one where one or two sides near the shadow are incorrect.

It is possible to calculate the roof boundaries and corners that cast the shadow; the corner formed by these roof-sides is likely to be very inaccurate, while the corner formed by the other two sides (non-shadow casting) is expected to be rather precise (otherwise no hypothesis would have been established). The gray-level along the sides of the roof is expected to change only on the non-shadow sides. The overall average gray-level should be low and the variance rather small.

This analysis leads to an easily derivable set of parameters that are used for the calculation of the most likely hypothesis. Figure 3.18 shows an example of a dark building not found by the automatic system. The line segments and junctions detected by the automatic system are shown in Figure 3.18 (b) and all the roof hypotheses are shown in Figure 3.18 (c). After specifying the detection problem, the image contrast is enhanced as shown in Figure 3.18 (d) for the display. A most likely hypothesis is selected in Figure 3.18 (e) and the error-ellipses of corners and center of gravity are shown. The error-ellipse indicates the uncertainty of a corner. The 3-D building model is found in Figure 3.18 (f) by just the initial interaction.

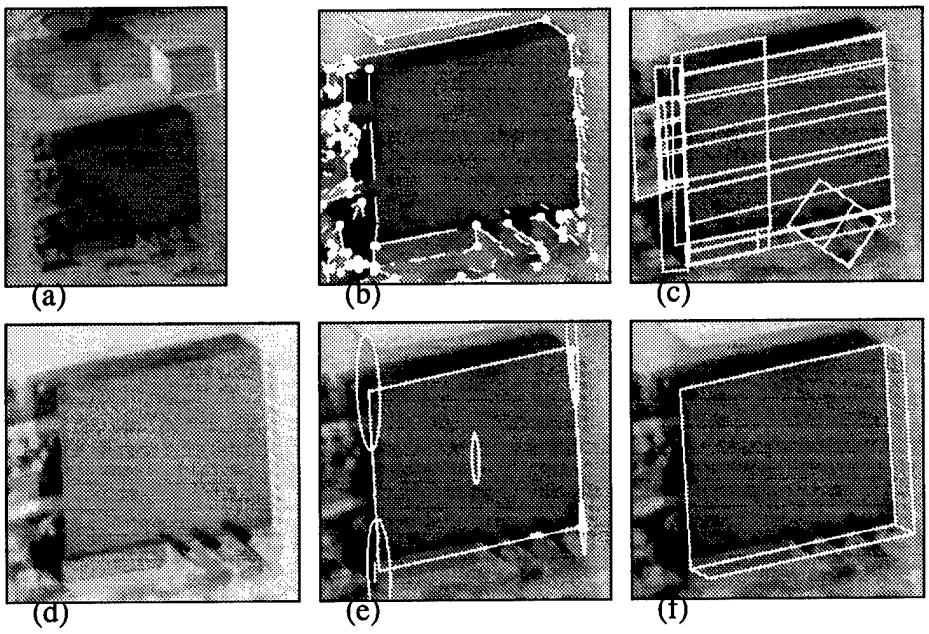


Figure 3.18 An example of a dark building recovered by the initial interaction

3.4.2 Corrective (Quantitative) Interaction

If the building is still not correctly detected, additional information is needed. One must back up one step in the hierarchy of the automatic system to extract new features, like edges or corners. Two ways of correcting the hypothesis are offered: first the user can adjust the roof-parallelogram by dragging sides with the mouse, rotating or translating the whole model. Changes can only be made within the constraints of the building model, for example, opposite sides remain parallel (see Figure 3.19). The extraction of a ground corner or edge (shadow corner or edge) will determine the building height. These interactions are similar to a completely manual system.

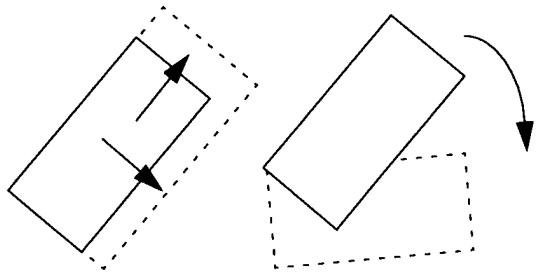


Figure 3.19 Manual adjustments - sides and rotation

Second, one can choose to extract edges and corners and associate them to a part of the building model. For example, a roof-side of the building can be specified by an edge extracted in the image. Then this edge is added to the current hypothesis (by replacing the nearest edge of the current hypothesis). Our system is implemented to run under the RCDE [6]. This environment allows the use of mouse-sensitive features that facilitate user selection and manipulation of features.

After each corrective interaction, the system forms a new parallelogram-hypothesis, looks for new edges, shadow, and wall evidence to support the new hypothesis, and finally performs a fitting and verification. These methods are the same as those in the automatic system. This important step of verifying the consistency of the constraints used in the automatic system is comparable to a fitting process in computer-assisted manual systems, though in our system, a fitting is performed after each interaction. Therefore it is possible that, after a manual correction of a roof-boundary, the wrong building height also is corrected automatically.

Without the fitting step the system would perform like a manual system and at least three interaction steps (two corner adjustments and one correction of the building height) would be necessary for adjusting the shape of one building model. Rotation and translation parameters may add another two steps. The manual feature extraction and the following fitting and verification steps can be applied to buildings that are automatically detected, but are partially wrong.

In Figure 3.20 the building is not detected because of missing edges. All roof-hypotheses generated by the system are shown in Figure 3.20 (c). There is no hypothesis correctly matched with the roof. After the initial interaction, a partly wrong roof-hypothesis, shown in Figure 3.20 (d), is found, where the shadow casting roof boundary is missing. The dotted lines show the estimated shadow boundary. The adjustment of one corner, shown in Figure 3.20 (e), leads to a new hypothesis. Note that after the correction of the corner, the system automatically finds the associated shadow boundary (dotted line) and it corrects the building height as shown in Figure 3.20 (f).

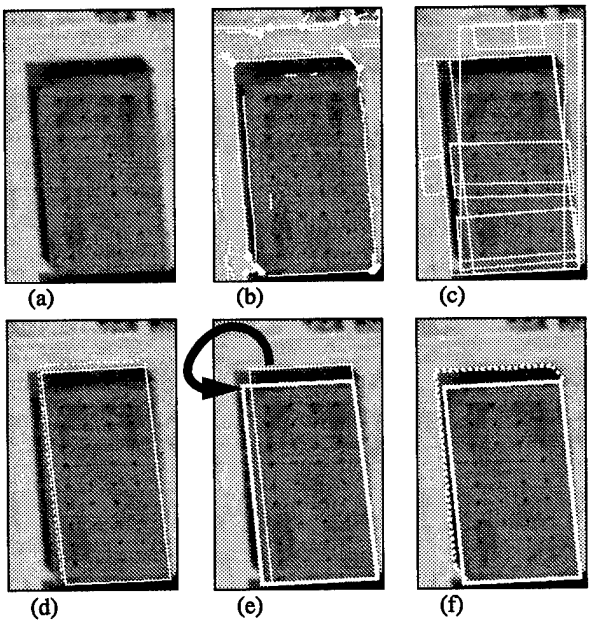


Figure 3.20 An example of corrective interaction

3.5 Results and Evaluation

This system has been tested on a number of examples provided by the RADIUS program with encouraging results. First, a few examples of the results of the automatic system are shown to demonstrate the performance of the system, and some of the sources of problems will be discussed in Section 3.5.1. A

partial evaluation of the automatic system also given in this section. Next, the results of the interactive system are presented in Section 3.5.3, followed by an evaluation of the interactive system.

3.5.1 Results and Evaluation of the Automatic System

This system has been tested on many examples from different sites. The examples shown here are all from the Fort Hood images provided by the RADIUS program. These images are taken from a real scene and present many difficulties for a typical suburban aerial image, such as vegetation, and complex shape buildings. An evaluation of the detection rate and the accuracy of the system is presented at the end of this section.

3.5.1.1 Examples

Figure 3.21 shows the results of several examples from the Fort Hood images. Figure 3.21 (a) and Figure 3.21 (b) show two L-shape buildings in Fort Hood image (fhn713). Note that parts of the shadows fall on the nearby vehicles. Although this makes the shadow boundaries highly fragmented, the system still successfully locates the correct shadow boundaries. Also, in Figure 3.21 (b) the building is dark and the wall on the left side of the building is inside the shadow and invisible. In this case, the building is verified by the strong shadow evidence.

In Figure 3.21 (c) and Figure 3.21 (d), note that there are some rectangular shape surface markings on the ground. The system actually makes roof hypotheses out of these surface markings. However, the system rejects these false hypotheses because no shadow or wall evidence can be found around them. Some parts of the building in Figure 3.21 (c) occlude the shadow and wall of the other parts, making the detection of such building more difficult. Such occlusions can be predicated from a partial analysis of the building; however, this system does not have this capability yet.

The building in Figure 3.21 (e) is composed of three parts. The part on the lower right corner has a different height from the other two parts. Although this system makes a hypothesis corresponding to this part, it can barely find wall or shadow evidence to support the hypothesis; it is difficult for people to determine the height of this part as well (the height would be easier to infer in an oblique view if a vertical side was visible). There are four gabled-roof buildings in Figure 3.21 (f). This system does not currently model gabled roofs; however, these examples are from a nadir view, hence it is able to detect three of these correctly. The fourth building, on the upper right corner, also is detected, but the description is slightly wrong.

Figure 3.22 shows the result of the automatic system on a window of the Fort Hood image (fhov1027). The system processes four regions inside this image window. The following discussion will take a closer look at the two areas on the right hand side of the image.

Figure 3.23 shows the details of the area on the lower right corner of the image in Figure 3.22. The system forms 1,204 hypotheses and selects 84 of them. Of these, 16 are verified. There are 14 buildings inside this window and 12 of them are detected. No false alarm is generated in this example. The system also has accurate descriptions on most of the detected buildings. There are two buildings not detected by the system. The one on the lower left corner of this window is a low building. The system does make a hypothesis of this building, however the shadow is not clear and the wall boundary is almost invisible. Therefore, the system does not have enough confidence on this building, and the building is not verified. Another building not detected by the system is the C-shape building in the middle of the window. This is a particularly difficult case because of its shape. The system made two hypotheses on the two wings of the building, but the evidence was not strong enough to verify any of them. The four L-shape structures attached to the four buildings on the left side of this window are not correctly described because shadows fall

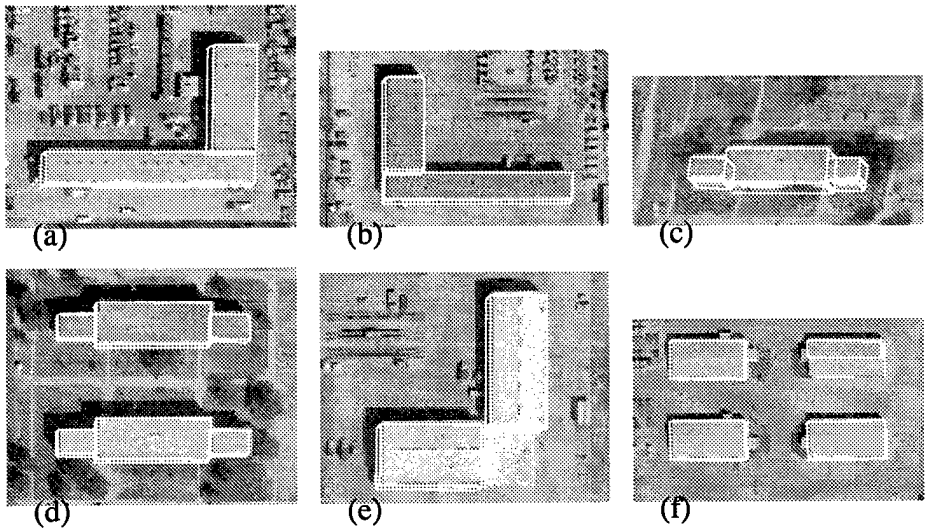


Figure 3.21 Fort Hood examples

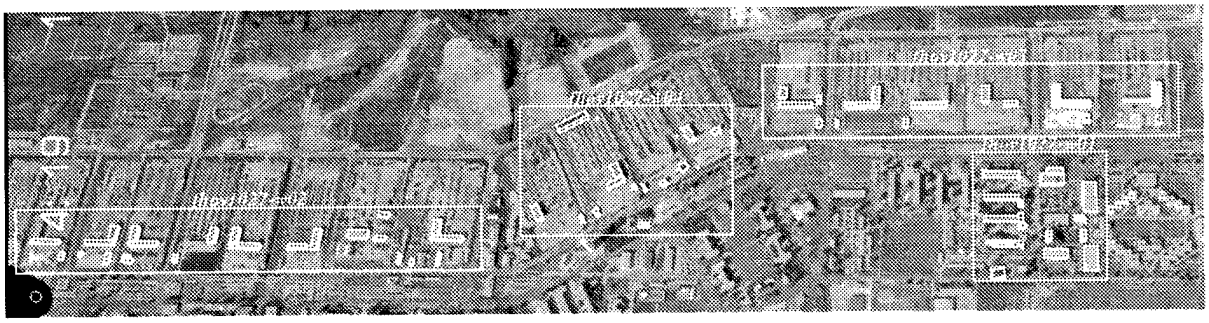


Figure 3.22 Automatic results for a Fort Hood image window

on them, their shadows are not clear, and the portions of visible walls are very small. The size of this image window is 670x645 pixels and the processing time is 686 seconds on a SUN SPARCstation 10. An evaluation of the detection rate and the accuracy of the detection on this example is shown in the next section.

Figure 3.24 shows the details of another area in the upper right corner of the image in Figure 3.22. In this example, the system forms 2,555 hypotheses, selects 102 hypotheses and verifies 16 hypotheses. There are 14 buildings in this area. The system detects 11 of them and verifies one false alarm. The false alarm is located at the right side of the left most L-shape building. It comes from a truck next to a dark region and it is very similar to a small building.

The system uses several parameters in the generation, selection, and verification of hypotheses. Some parameters, such as the search range of wall and shadow evidence, can be set as a function of the image resolution. Some parameters, such as the weights used in the wall and shadow evaluation functions, are chosen based on our experiences on several test examples. We also can have a learning program to find the best parameters over a set of training examples. All the results shown here use the same parameters.

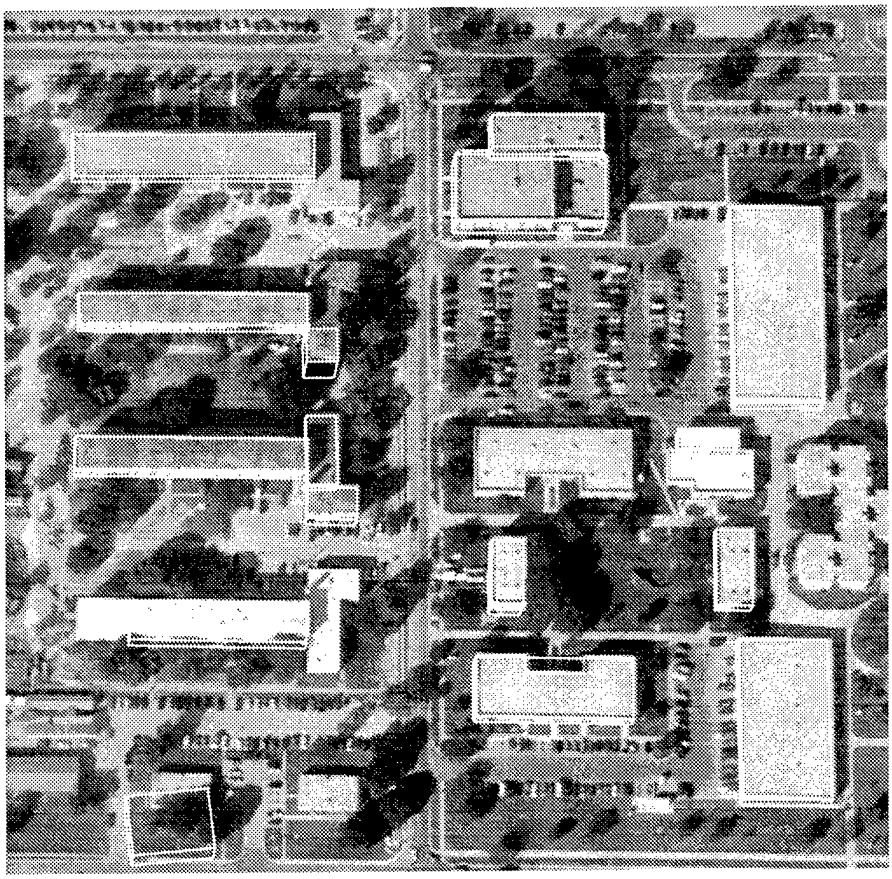


Figure 3.23 Automatic results for a portion of a Fort Hood image

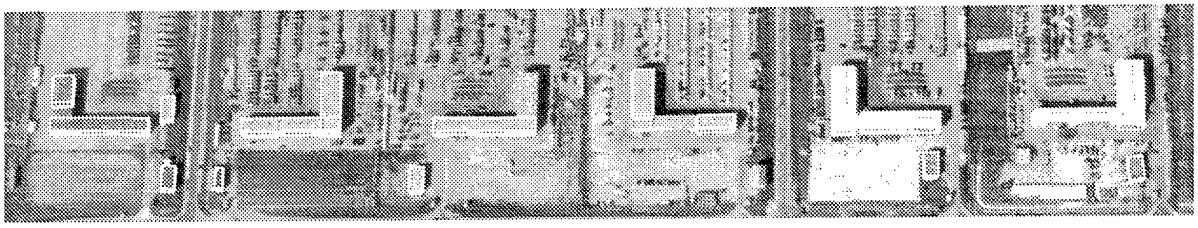


Figure 3.24 Automatic results for another portion of a Fort Hood Image

3.5.2 Detection Evaluation

There are many ways to measure the quality of the results [24,28]. We use the following five measurements (explained in the next paragraph) for evaluation:

- Detection Percentage = 100 x TP / (TP + TN)
- Branch Factor = 100 x FP / (TP + FP)

- Correct Building Pixels Percentage
- Incorrect Building Pixels Percentage
- Correct Non-Building Pixels Percentage.

The first two measurements are calculated by making a comparison of the manually detected buildings and the automated results, where TP (True Positive) is a building detected by both a person and the program, FP (False Positive) is a building detected by the program and not by a person, and TN (True Negative) is a building detected by a person and not by the program. The other three measurements are calculated by labeling every pixel in the image as either a building pixel or a non-building pixel [24, 28]. We calculate the percentage of the number of pixels correctly labeled as building pixels over the number of building pixels in the image, the percentage of the number of pixels incorrectly labeled as building pixels over the number of pixels labeled as building pixels, and the percentage of the number of pixels correctly labeled as non-building pixels over the number of non-building pixels in the image. Table 4. shows the evaluation on the results of our system on four image windows shown in Figure 3.22.

Table 4. Detection Evaluation

	Detection Percentage tp/(tp+tn)	**Branch Factor fp/(tp+fp)**	**Correct Building Pixels**	**Incorrect Building Pixels**	**Correct Non-Building Pixels**
fhov1027-w01	85.7%	0.00%	78.8%	8.71%	98.6%
fhov1027-w02	72.0%	0.00%	74.4%	1.45%	99.9%
fhov1027-w03	78.6%	8.33%	78.4%	0.81%	99.9%
fhov1027-w04	64.3%	18.18%	42.7%	29.71%	99.1%
Average	74.6%	5.66%	71.7%	7.40%	99.5%

A building is detected if a part of the building is detected by the system. The description of the detected building might not be correct. The evaluation of counting correct building and non-building pixels gives us an approximate idea of how accurate the description is. Note that our system gives rather consistent results for most images, except for fhov1027-w04, an area where the orientation of the L-shape buildings in the image is almost parallel to the direction of illumination and the other orientation of the L-shape buildings is almost parallel to the projection of the vertical line. Therefore, only one side of the roof casts a shadow, and only one side of the walls is visible; it is difficult for the verification process to find enough evidence to verify these L-shape buildings. Also note that the gray level of three of the L-shaped buildings is very similar to the gray level of the surrounding ground and this makes the roof boundaries of these buildings very fragmented.

3.5.2.1 Confidence Evaluation

Our system associates a confidence value with each hypothesis that can further be used to evaluate the performance of the system and guide a user on how to interpret the results. Figure 3.25 shows a histogram of the number of true and false positives of the results in Section 3.5.1.1 corresponding to certain con-

fidence levels (ranging between 50 and 100, in increments of 5). The confidence values which are between 0 and 1 have been scaled to the range of 0 and 100 for display purpose.

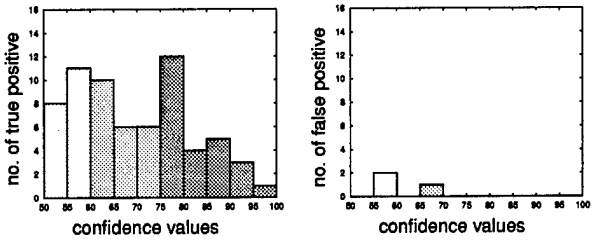


Figure 3.25 Distribution of confidence values

Note that there are only three false positives, and they all have low confidence values. In fact, if we set a confidence threshold of 70, we detect no false positives at all and that more than half of the true positives are above this threshold. This indicates that the confidence values can be used profitably by an end-user or by another program. Results given with high confidence can be taken to be reliable, and further attention for improving the results can focus on the lower confidence results, if necessary. We believe that this self-evaluation capability will greatly ease the use of our automatic tool in an interactive environment.

3.5.3 Results and Evaluation of the Interactive System

In this section we present the results and evaluation of the interactive system on the same examples used by the automatic system in Section 3.5.1. First, we show the results, and then an evaluation, on the number of interactions is given.

3.5.3.1 Examples

Figure 3.26 shows the final results of the interactive system on the same image as in Figure 3.22. We use different colors (a color image is available) to show how many interactions is required for each building. A closer look and discussion of the results at the window (fhov1027-w01) on the lower right corner of this image is given in the next paragraph.

In Figure 3.27, we can see the results are very accurate within the limit of the image resolution; the system can accurately model a rectilinear building. Therefore, we are not going to discuss the accuracy of the results. Instead, we focus on the number of interactions required to generate the results. Here the basic component of the building model is a cube object. An L-shaped building is composed of two cube objects and a C-shaped building can be modeled as a combination of three cube objects as shown in the figure. A building can be composed of several cube objects with or without the same height. We call the cube object a structure component of a building. The interactive system modifies the structure components of each building, if necessary. Our evaluation of the interactive system calculates the number of interactions required on the structure components to correctly model all the buildings in the scene. .

3.5.3.2 Evaluation

The evaluation of the system for the examples of Section 3.5.3.1 is shown in Table 5.. Note that initial interaction suffices to correct the problems in several cases. In nearly all of the cases where corrective interactions are required, only corrections of the sides and height are necessary, because rotation and position are already given by the hypothesis selected by the initial interaction. Also, most of the number of corrective (quantitative) interactions required are less than three.

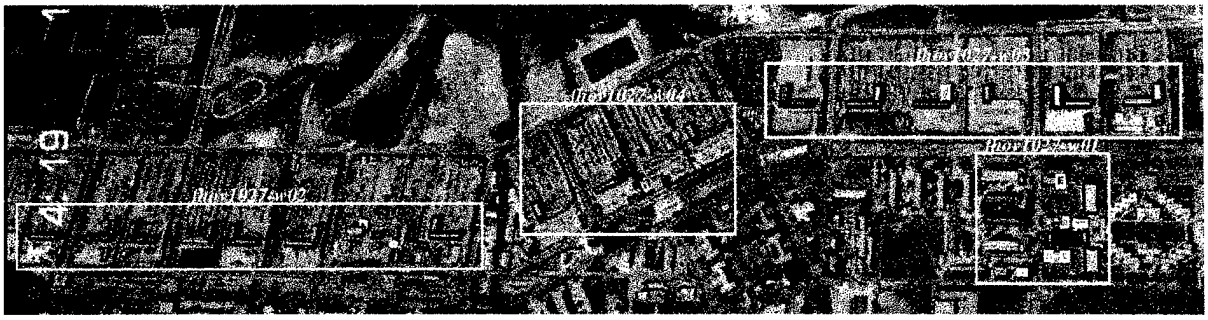


Figure 3.26 Semi-Automatic results for a Fort Hood image window

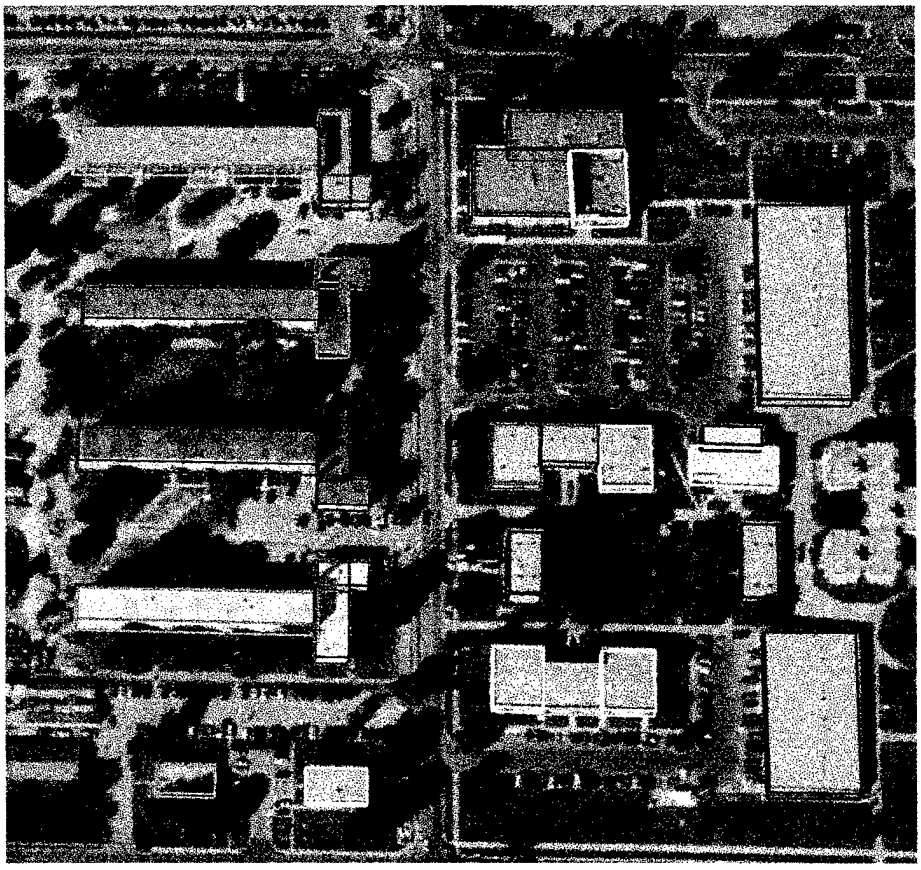


Figure 3.27 A Result of the Interactive System on a Fort Hood image

Sixty two and a half percent of the cube object components are detected and they required 37 side or height corrections. The rest of the cube object components of these buildings are found by the initial interaction and a total of 50 corrections.

The total number of interactions required to generate the results in Figure 3.26 is 126, including 39 initial interactions and 87 corrective interactions; there are 104 cube objects in the image. A manual system

will need at least 416 interactions to generate the same results (each cube object requires 4 interactions). This shows how dramatically our system can aid in creating the site model.

3.6 Conclusions

We described a system for detection and description of buildings from a single aerial image.

The building height estimation is more reliable with the use of both wall and shadow evidence. We believe that the results show that the system gives good performance, particularly on large buildings with reasonable contrast and shadows. We also believe that the confidence measures offer a tool that can help utilize the results even when they are not perfect. We intend to work on extending the range of imaging conditions and complexities of shapes that our system can handle. We also will add a reasoning process to analyze the final results to give the system feedback that can aid it in the detection of some of the missing buildings.

Table 5. Interaction Evaluation

Required Interaction	Color	Number of Structures	Cumulated Percentage	Number of Corrections
Detected with Correct Description	Red	38	36.54%	0
Detected + 1 Correction	Magenta	20	55.77%	20
Detected + 2 Corrections	Salmon	4	59.62%	8
Detected + 3 Corrections	Orange	3	62.50%	9
Not Detected + 1 Qualitative Correction	Yellow	10	72.12%	10
Not Detected + 1 Qualitative + 1 Quantitative	Light Green	13	84.62%	26
Not Detected + 1 Qualitative + 2 Quantitative	Green	11	95.12%	33
Not Detected + 1 Qualitative + 3 Quantitative	Cyan	5	100.0%	20
TOTAL		**104**		**126**

4 Automatic Model Construction from Multiple Views

The task of building detection and description from aerial images is a difficult problem owing to many factors. Outdoor environments are not controlled environments, giving rise to viewpoints and illumination conditions that may vary considerably. The sites themselves may be fairly complex, containing a number of elements, such as trees and vegetation, that add to the features detected, as well as parking lots, fencing and roads, which are geometrically similar to buildings. Segmenting buildings is a non-trivial task owing to the fragmentation that is generally observed.

It is possible to recover the desired building structures from a single image and some automatic systems have been constructed for this task [9, 24]. However, this is an extremely difficult task as only one view is available to resolve the ambiguities of segmentation, and 3-D recovery must rely on shadows and projected lengths of vertical lines which may not be visible distinctly. For many applications, more than one view of the scene is available, which can simplify the task significantly. This report considers the case where two or more views are available. The views are not necessarily taken at the same time; hence, imaging conditions, including the sun position, the atmospheric conditions, and the environmental conditions, may be quite different.

Problems of segmentation and 3-D recovery are simplified by the presence of multiple views, but do not disappear completely. A simplistic view of multiple view processing would be that a dense 3-D map might first be recovered by matching across the different views followed by segmenting the desired structures in 3-D. However, this is rarely possible in stereo processing and is particularly difficult for the problem being considered here. It is not possible to directly compute a dense 3-D map of the scene as there are large homogeneous areas whose interiors cannot be matched directly, and therefore it is not possible to match intensity values across images because the values are not invariant with changing viewing conditions. Instead, what is attempted is matching features, such as object boundaries, that are invariant across the images. Because the set of features will likely be sparse and fragmented, they must be grouped [29] to infer coherent objects.

To illustrate the nature of the problem, consider three views of a scene shown in Figure 4.1. These views come from a modelboard, and are being used as standard test images by several researchers. Note that the sides of the buildings that are visible are not the same in all views, and that the shadows cast on the ground are quite different. Figure 4.2 shows the lines extracted from the images in Figure 4.1 images. Note that not all of these boundaries have correspondences in more than one view. Also, determination of unambiguous matches is unlikely even for those lines that do correspond just by looking at the lines individually, as many parallel lines are likely to be present nearby in an urban scene where buildings are often parallel to each other, as are ancillary structures, such as roads, sidewalks and landscaping.

There have been previous attempts focused on detecting buildings from stereo images [15, 16, 25]. These systems assume that images are acquired during a single session, typically the same day, and under

similar illumination and environmental conditions. For the most part, previous systems match low-level features, such as line and junctions, and attempt to infer buildings from the matches by some kind of tracing or grouping method. The system described in [15] matches higher level hypotheses (rectangles) but does not use stereo information to form the hypotheses themselves. One recent system does deal with RADIUS imagery taken at different times [30], and a comparison with this system is presented later in this section.

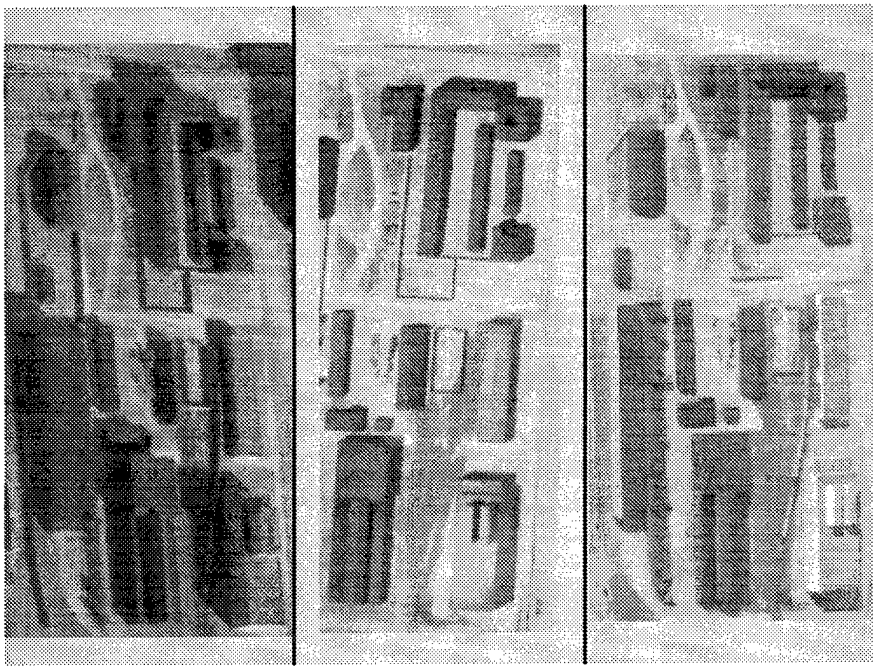


Figure 4.1 Three views of a scene

For the detection and description of buildings from multiple views, it is suggested that the problems of matching and grouping (i.e. 3-D recovery and object segmentation) not be separated, but be solved simultaneously [31]. The difficulty with matching lower-level features is that it is difficult to disambiguate the matches correctly; the difficulty at the higher levels is that the correct groupings may not be formed in the first place. A hierarchical grouping scheme where lower level features, such as lines, junctions, parallels and U structures, are grouped into successively higher-level features, is proposed. At each level, the grouped structures are matched across the different views and only the consistent ones are retained. The initial thrust is to first recover rectangular roof structures, which are parallelograms in each view projection, because they form the dominant regions of the buildings in the projected images. In addition, selection of roof hypotheses must take advantage of the context provided by the visible walls (which may be different in different views) and by shadows cast by them.

The use of features in building detection and description is appealing because features are usually view independent. However, in general, edge detection (or general feature detection for that matter) does not produce perfect features in general. Perceptual grouping is used to relate features that could possibly be fragments of larger features. Multiple views enable matching of features at virtually all levels. By matching grouped features at several levels in the hierarchy described in Figure 4.3, the system becomes more robust to incomplete information, and has the desirable property of graceful degradation with degraded input data.

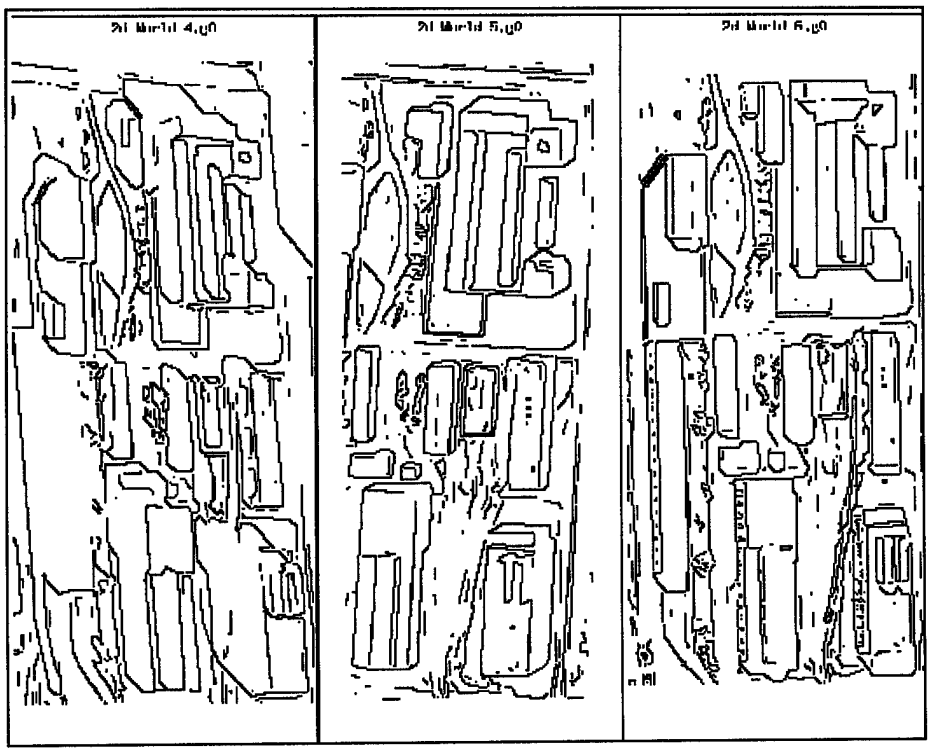


Figure 4.2 Detected line segments for Figure 4.1

To simplify the task, the domain of buildings is restricted to rectilinear structures (i.e. those consisting of rectangular components). Further, it is assumed that the roofs are planar and that the walls are vertical. This allows some predictions about the expected properties of the projected boundaries in the image as explained later in Section 4.2. It also is assumed that the "camera models" are given, i.e., it is possible to infer the epipolar geometry between the views and know the orientation with respect to a ground frame. There is no requirement that the different views be such that the epipolar lines are parallel, nor is there an attempt to *rectify* the images to parallelize the epipolar lines. The system described in [30] uses a single view to generate rooftop hypotheses, and verifies them using the other available views. The system detailed in this chapter uses all the available views to generate hypotheses, and verifies them based on the accumulated evidence from the views.

Our system uses several parameters in its decision-making processes, at various levels. While the choice of parameter values determines the quality of the results that are obtained, attempts have been made to desensitize the system to small variations in these values, as far as possible. Ideally, the parameter values should be based on a mathematical analysis of the algorithms and be a function of the parameters of the input images and any known parameters of the site. Unfortunately, such an analysis is difficult because of the complexities of the algorithms and estimating appropriate image parameters. In the current system, decision parameters are a function of the image parameters that are supplied with the images, such as the resolution.

The parameter values have been set by an informal analysis of the process, and testing with limited data. All the examples use the same parameter values. It is not clear that these values also will work well for other images and other sites. Experience indicates that new images and new sites do not so much require

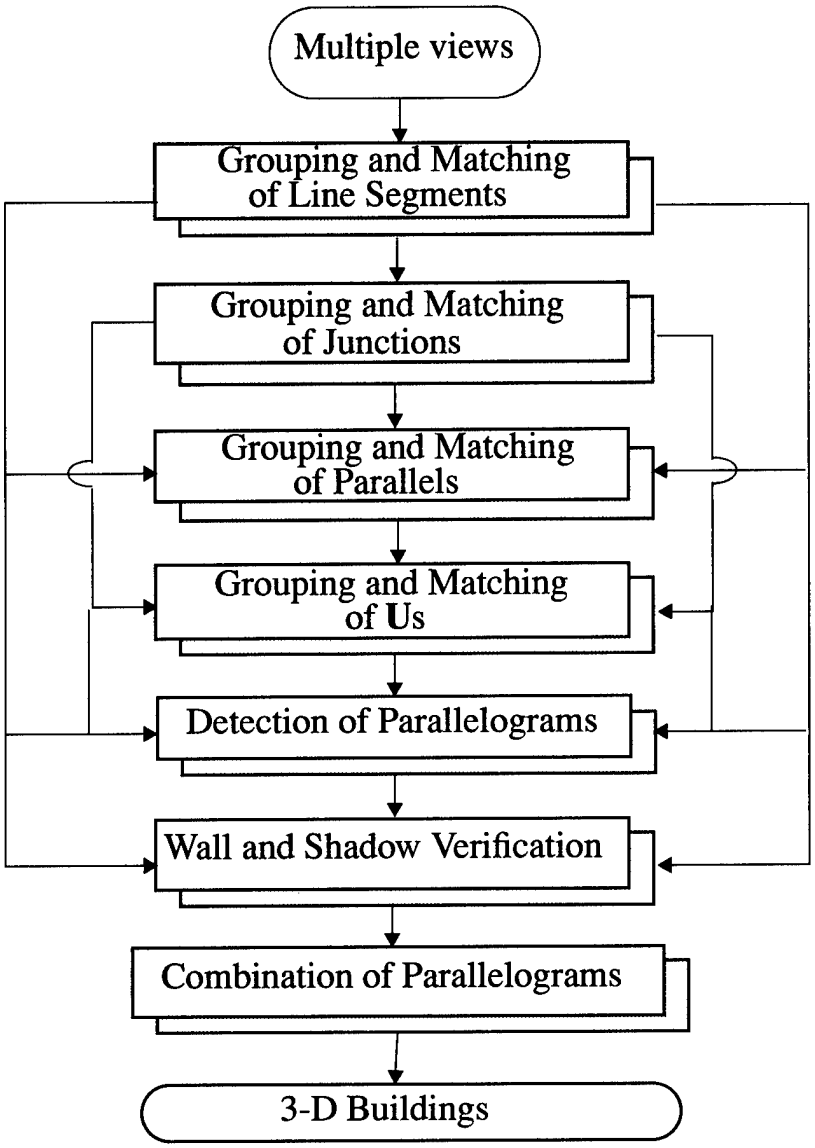


Figure 4.3 Block diagram of the system

changing of the parameters, but perhaps there is a need for additional reasoning steps to deal with situations not encountered in earlier tests.

Details of this system are presented in Sections 4.1 through 4.6. Results and conclusions are presented in Section 4.7.

4.1 Hierarchical Grouping and Matching of Features

This section presents detailed descriptions of the features used in the current system, including the methods for grouping and matching them. As described above, the system is hierarchical and uses evidence from all the views in a non-preferential, order-independent way.

4.1.1 Lines

Before matching, colinear line segments are grouped. Segments are considered colinear if there is a *free path* from the end of one segment to the other, i.e. no other segment blocks, if there is a line joining the two closest endpoints of the colinear segments, and if the angle between the segments is less than 10 degrees. Colinearity is applicable to a set of greater than two segments as well. The above criterion must be met between every pair of neighboring segments.

After colinear grouping, the lines are tested for matches across the views by using the following quadrilateral constraint:

- The match for a line segment in one view must lie at least partially within a quadrilateral owing to epipolar and 3-D height constraints.

Each pair of lines that meet the quadrilateral constraint in any pair of views is determined to form a *line match, is* included in the set of line matches that is denoted by S_{lm}, and is passed to the higher levels for further processing. The complexity of the line matching algorithm for a single line is proportional to its length, hence the process is linear in the summed lengths of the lines.

4.1.2 Junctions

Next, matching of junctions formed by the intersection of two lines is considered. Consider a pair of lines L_{ik} (k = m, n) in $view_i$, with endpoints P_{ikl} (l = 1, 2). Junction J_{ij} is formed at the intersection of L_{ik} (k = m, n) iff the angle between L_{im} and L_{in} is greater than 30^o and min (distance(J_{ij}, P_{ik1}), distance(J_{ij}, P_{ik2})) $\leq$ length (L_{ik}) for (k = m, n). Denote the set of junctions formed in $view_i$ by S_{J_i}. Junctions in the sets S_{J_i} (i = 1, 2... number_of_views) are then matched across the views to form a set S_{jm}, when the following constraints are satisfied:

4.1.2.1 Epipolar Constraint

Given a junction J_{ij} in $view_i$, its match in another $view_l$ must be within a certain segment (depending on the height range in 3-D) of the epipolar line corresponding to J_{ij}, in $view_l$.

4.1.2.2 Line Match Constraint

If junction J_{ij}, formed by lines L_{im} and L_{in}, matches junction J_{kl}, formed by lines L_{kp} and L_{kq}, then exactly one of the following must hold: either there exist line matches (L_{im}, L_{kp}) and (L_{in}, L_{kq}) in S_{lm}, or there exist line matches (L_{im}, L_{kq}) and (L_{in}, L_{kp}) in S_{lm}.

4.1.2.3 3-D Orthogonality Constraint

Given a junction match, it is possible to compute the 3-D angle between the lines forming it (from the knowledge of the matching lines). It is required that this angle be between 80 and 100 degrees in 3-D.

4.1.2.4 Trinocular Constraint

When there are more than 2 views available, the well-known trinocular constraint may be applied to the locations of the junctions.

Junction matching is linear in the number of junctions formed in all the images. The formation of junctions in each image is $O(n^2)$, where n is the number of lines, in the worst case. However, since the formation of junctions is limited by the length of the lines, it has been observed that the number of junctions

is O(n). In addition, the formation process uses the constraint due to the length of the lines, hence, the process too, is O(n).

4.1.3 Parallels

Next, *parallel* pairs of lines and matching of the parallel pairs are computed. Parallels are formed between pairs of lines, L_{ij} and L_{ik} in the same $view_i$, that are separated by less than the maximum projected width of a building. While the task domain causes a large number of parallels in each view (two to three times the number of lines in that view), because of the alignment of buildings, roads, parking lots and shadows, the number of parallel matches is typically lower than the number of lines in any view. A match is hypothesized if there is evidence in at least two views. When there is evidence in greater than two views, this forms a single parallel match in more than two views. The constraint used in matching is the parallel match constraint described in the following:

4.1.3.1 Parallel Match Constraint

Consider parallels P_{ik} with component segments L_{ik_1} and L_{ik_2} in $view_i$, and P_{jl} with component segments L_{jl_1} and L_{jl_2} in $view_j$. The parallel match constraint is satisfied for this pair of parallels if and only if exactly one of the following criteria is met:

- (L_{ik_1}, L_{jl_1}) and (L_{ik_2}, L_{jl_2}) are elements of S_{lm}
- (L_{ik_2}, L_{jl_1}) and (L_{ik_1}, L_{jl_2}) are elements of S_{lm}

In the case of parallels over more than two views, the parallel match constraint must be satisfied over parallels from every pair of views. Maximal parallel matches, i.e. parallel matches that have the maximum number of parallels, are generated in order to ensure that duplicate parallel matches do not occur. Parallel matches over n views are represented as n-tuples. The set of parallel matches is denoted by S_{pm}. Note that as the line matches satisfy the quadrilateral constraint (they are constrained to being in a certain range in world z values) order reversal of the lines in the other views is automatically taken care of, if it should occur. The complexity of parallel matching is linear in the number of parallels in each image, as the search is spatially constrained by the epipolar constraints. The formation of parallels is implemented using hash tables, and is proportional to the number of detected lines weighted by their length, in each image.

4.1.4 Us

Next the formation of **U** structures is considered. A **U** captures 3 sides of a parallelogram. **U**s are formed when two junctions are *aligned*. The definition of alignment is given below. **U**s are computed for each $view_i$, to form sets S_{U_i} (i = 1, 2... number_of_views). These sets are used in forming parallelogram matches as detailed in section 4.1.5

There are two ways in which the system hypothesizes **U** matches. The first way of hypothesizing **U** matches is by using 2 *aligned* junction matches. Junction matches Jm_p and Jm_q are *aligned*, if their component junctions are *aligned* in each view. *Alignment* of junctions is illustrated in Figure 4.4.

The second way that **U** matches are formed is by a parallel match with evidence of *closure* in at least one view. In this case there are two virtual junction matches hypothesized. These two virtual junction matches are hypothesized on the side of the **U** match where evidence of *closure* exists. The intersection of this *closure*, in each view (it is hypothesized in views where it does not exist), with the component parallel of the parallel match in that view, yields the virtual junctions that are components of the virtual junction matches.

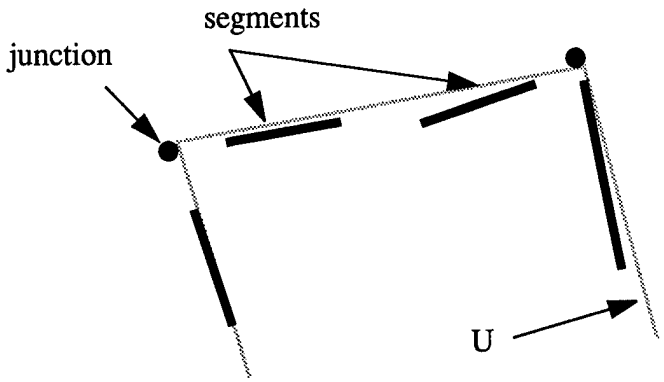


Figure 4.4 Aligned junctions

Closure of a parallel is defined as follows. Let the line that extends further in the parallel be L_{ij}. Let the further endpoint of L_{ij} be P_{ij1}. Let $L_{orth_{ij}}$ be the projection of the line in 3-D that is orthogonal to the line in 3-D that is parallel to the ground and projects to L_{ij}. The parallel is said to have closure if and only iff there exist segments which form an angle of less than 10 degrees with $L_{orth_{ij}}$, whose endpoints are at a distance of less than f(resolution) from $L_{orth_{ij}}$, and whose perpendicular projections cover at least 50 percent of the distance between the parallel lines along $L_{orth_{ij}}$.

Denote the set of **U** matches by S_{um}. In the first case of **U** match formation (from two aligned junction matches), the following constraint should also be satisfied:

4.1.4.1 Planarity Constraint

Each junction match defines a plane in 3-D. The planarity constraint checks that the planes of the junction matches forming a **U** match, are approximately coplanar (approximate coplanarity is defined to mean that the angle between the normals is less than 10 degrees, and the distance is less than f(resolution), in each view).

Formation of U matches is linear in the number of parallel matches that are formed, weighted by the distance between the parallels that form the parallel match in each image (because the search for U completions is performed in the space between the parallel lines).

4.1.5 Parallelograms

Formation of parallelogram matches is the basis for hypothesizing building roofs. To hypothesize a parallelogram match, the minimal requirement is a **U** match or a parallel match. Parallelogram matches are hypothesized from the available U match (or parallel match), with a search for completions for the U match (or parallel match). The existence of evidence to form a parallelogram match is a strong indication that a rectangular 3-D structure exists. Parallelogram matches across the views are the hypotheses of rectangular 3-D buildings.

4.2 Selection of Roof Hypotheses

The parallelogram matches serve as roof hypotheses. They satisfy the constraints of being rectangular in 3-D, and almost coplanar. In addition, the height and orientation with respect to the ground is

known. These parameters may be used to distinguish among acceptable hypotheses and others, but height and orientation constraints are not sufficient because of inaccuracies of feature detection and camera models, and the possibility of erroneous matches giving rise to acceptable hypotheses accidentally. In addition to the 3-D constraints that the roof hypothesis must satisfy, it must be acceptable in all the 2-D views. This gives rise to the second constraint of the three described below. The hypotheses retained are shown in Figure 4.5 below:

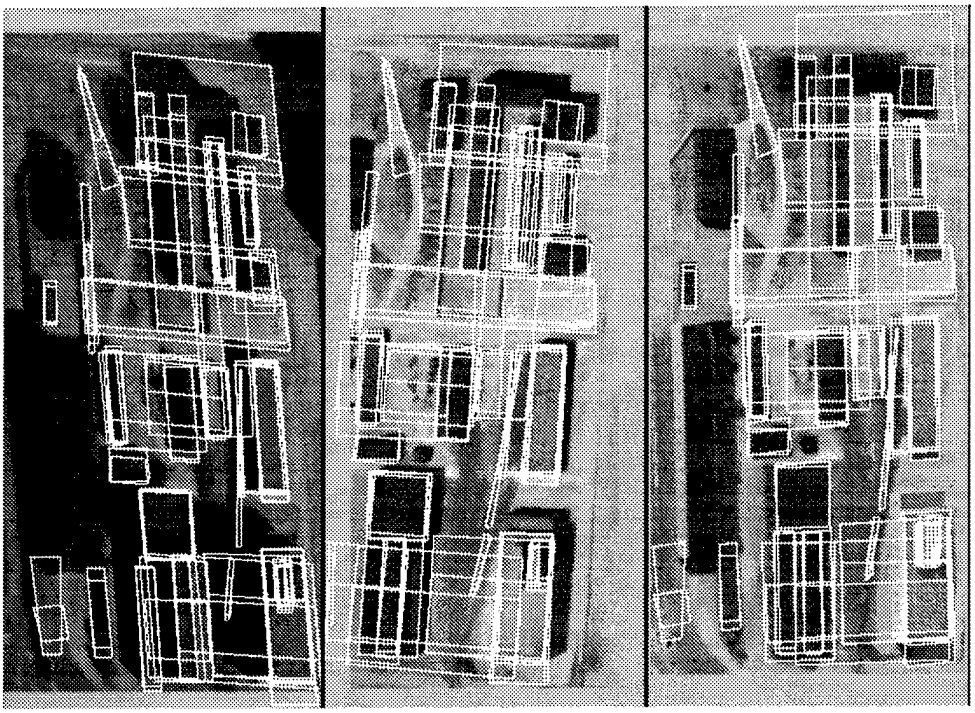


Figure 4.5 Selected parallelogram matches (building hypotheses)

4.2.1 3-D Height

A check on the height constraint is necessary at this stage, even though the features forming the parallelogram match satisfy the constraint.

4.2.2 Positive and Negative Line Evidence

If the height constraint is satisfied, a search is performed in each view for evidence supporting, or negating, the roof hypothesis. Lines that are found within a certain distance (which is a function of the image resolution) from the hypothesized parallelogram, and that differ in angle by not more than 10 degrees, are considered positive evidence. Negative evidence consists of lines that cross the boundaries of the hypothesis. Figure 4.6 illustrates the concept of positive and negative evidence.

4.2.3 Orientation

The normal of the plane containing the roof hypothesis must make an angle of 45 degrees or less, with the normal to the ground in 3-D, to be considered for verification. The components of the parallelogram match, making the roof hypothesis satisfy this constraint.

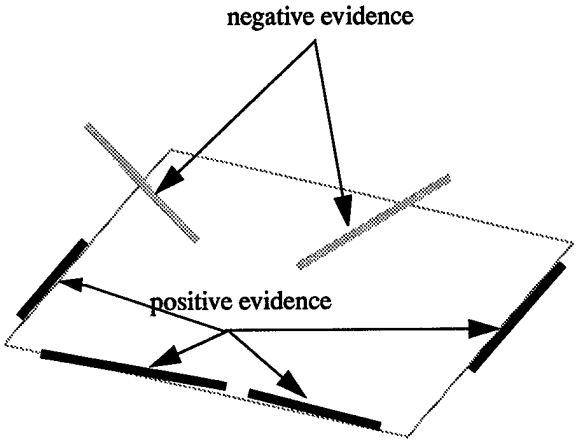


Figure 4.6 Positive and negative line evidence

4.3 Verification of Building Hypotheses

So far, the only evidence used was from the component features of a single roof hypotheses. If there is indeed a building roof, it should be possible to find other features that come from the 3-D nature of a building. In the system described here, there is a search for evidence for walls and that of shadows cast by a hypothesized roof. In addition to the evidence of features supporting or negating roof hypothesis, statistical properties of the regions of the hypothesized roof and the shadows cast are factored in.

4.3.1 Wall Evidence

In a view which is not nadir (and most views can be expected to be such), at least one and not more than two of the side walls of the buildings will be visible. These walls are assumed to be vertical. The verification for walls involves looking for the projections of the horizontal bottom of the wall (the interface of the vertical wall and the ground). At this point the height of the hypothesized building is known through triangulation. Using the camera models, the projection of the vertical direction in 3-D is computed. From the top of the wall to the bottom, a search for line evidence parallel to the side of the hypothesized building is performed in incremental steps. Figure 4.7 illustrates this concept. Wall evidence is deemed to be found if there is evidence of parallel lines at the distance from the top of the building that is predictable from its height in 3-D. The score associated with this evidence is a function of the ratio of the length of the line coverage of the side to the length of the side.

As additional evidence for the presence of a wall, a search for the projection of the visible vertical sides of the hypothesized building is performed. If the predicted length of the projected vertical (obtained from the height of the building in 3-D) is less than five pixels, it is considered unreliable, and not taken into consideration. Each vertical wall that is found increases the confidence of the hypothesis.

4.3.2 Shadow Evidence

A 3-D building structure should cast shadows under suitable imaging conditions. The system should normally possess knowledge of the direction of illumination from the sun, which in turn allows it to predict the location and orientation of shadows (on flat ground) from the 3-D hypotheses. If such shadows are found, the confidence in the hypothesis can be increased. Shadows have previously been used in monocular

detection of buildings [9]. In the present case, the analysis is simpler because there is an estimate of the height of the building *before* there is the need to search for shadows.

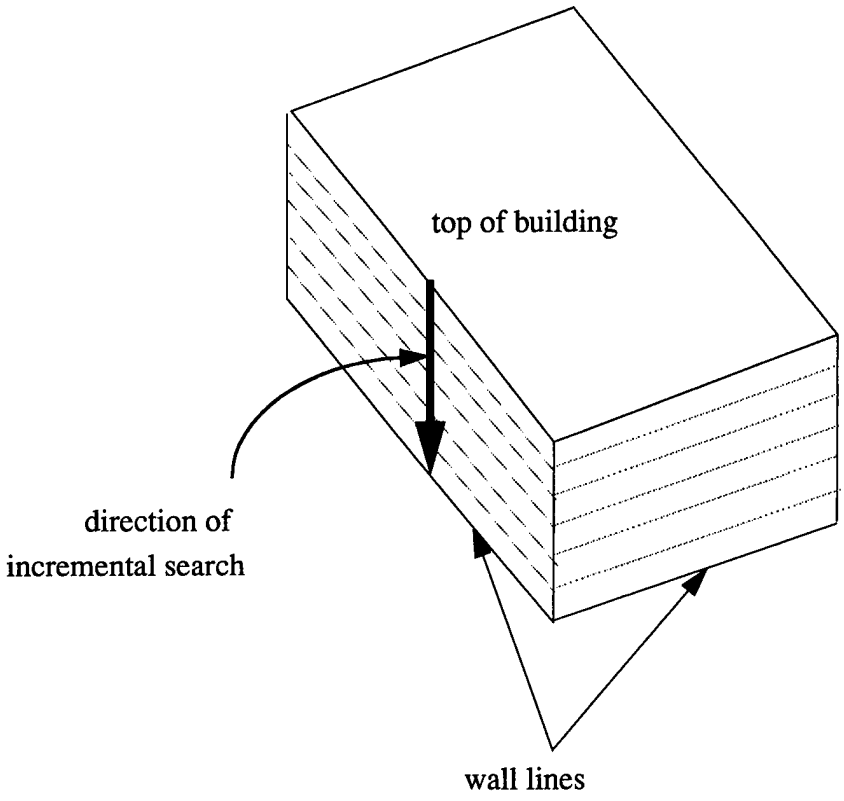


Figure 4.7 Search for walls

The search for shadows is carried out in a manner similar to that for the walls. Knowing the height of the building in 3-D, and the direction of illumination, a search is performed to detect evidence of the predicted projection of the shadow. This includes the shadow cast by the detected roof, and the shadow cast by the vertical walls of the building. Occlusion of shadows by the building itself is taken into consideration when searching for shadows. The search for shadows in each view is carried out separately because the views are obtained with different sun positions. Hence shadows are strictly monocular cues.

The visible sides of the walls are dependent only on the 3-D orientation of the building relative to the camera. The sides of the building for which the shadows are visible is dependent on the orientation of the building, and the direction of illumination. As these parameters (namely the viewpoint and the direction of illumination) are independent, it is possible that the shadows cast by the roof, and the vertical wall of the same side, are visible on the same side of the building. In this case, the search is performed simultaneously. The shadow lines and the wall lines may be visible together, depending on whether the material of the building and the diffused light allow detection of the wall lines, which lie in the shadow area in this case.

The evidence of shadows and walls is accumulated from all of the views, and a score is associated with the evidence detected. This score is a function of the extent of coverage, against expected coverage, and the accuracy of the location of the evidence compared to the predicted location. However, the system does not take into account missing evidence because of occlusion by other detected structures, or because

of the shape of the building. For instance, if the structure is L-shaped, the system might hypothesize the structure as a combination of two adjacent or two overlapping rectangles. In either case the two rectangular hypotheses may lack evidence in the common part, depending on the viewpoint and on the direction of illumination. The numerical evidence for walls and shadows is compared against a threshold, T, for acceptance or rejection. The threshold, T, is set empirically from tests on sets of data from different scenes. Mathematically, if the wall evidence of a building, B_j in $view_i$ is $wall_{ij}$, and its shadows evidence is $shadow_{ij}$, then B_{ij} is verified if and only if:

$$\sum_i (wall_{ij} + shadow_{ij}) \geq T$$

Figure 4.8 depicts the results obtained from the set of views shown in Figure 4.1; 13 of the 16 buildings have been detected. Some of the detected buildings have markings on the roof, and one is a compound building with a rectangular outline. A compound building in the lower right corner has the top (and dominant) level detected. The second (lower) level has an area that is smaller than the smallest expected buildings. The buildings that are not detected are dark, or those whose boundaries merge with their shadows. This makes detection difficult as U matches and parallel matches are not formed.

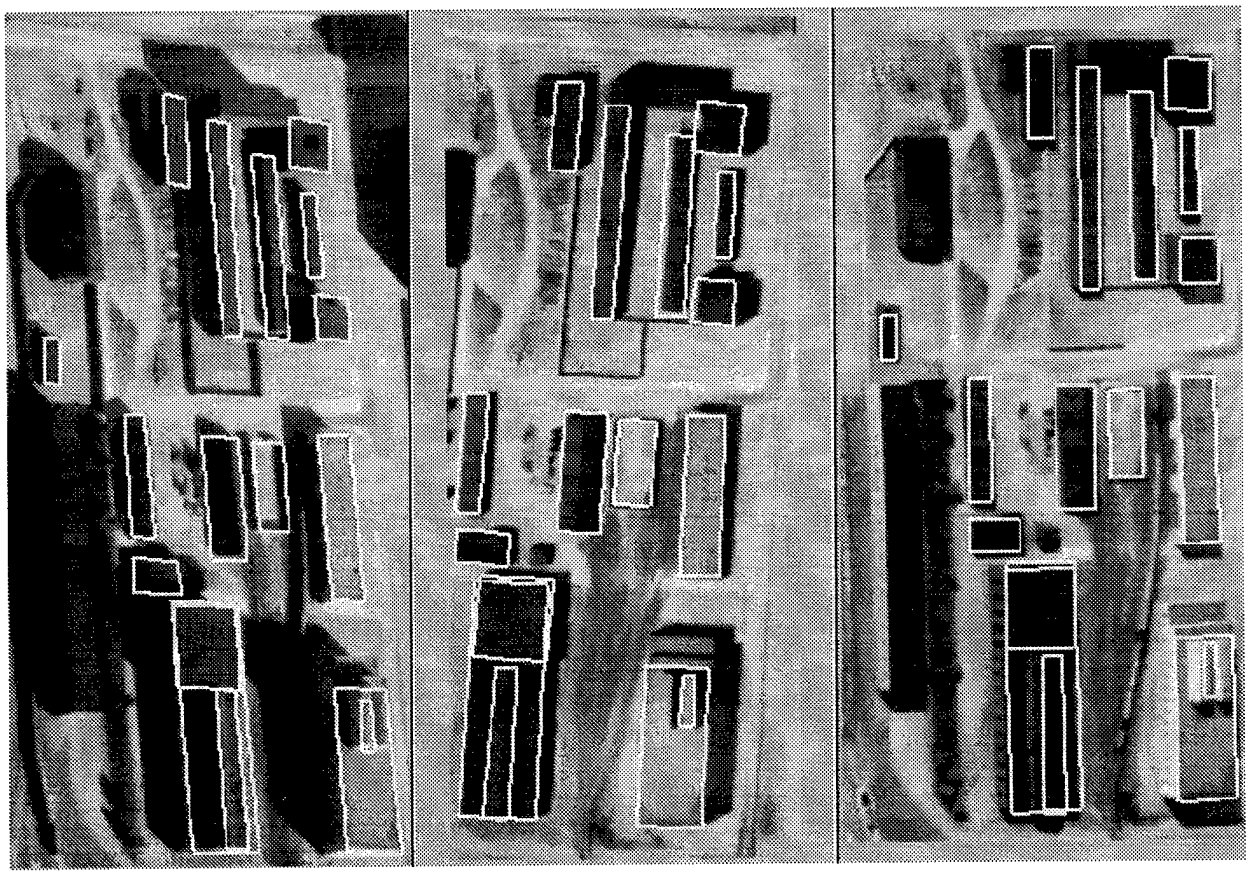


Figure 4.8 Results on the views in Figure 4.1

4.4 Combination of Rectangular Buildings

Some buildings are not rectangular, but can be decomposed into rectangular structures. Verified rectangular hypotheses are examined for combination according to two mutually exclusive criteria: proximity, and overlap. The precondition for both criteria is that the hypotheses be of approximately the same height in 3-D.

Proximity: When two hypotheses have common boundaries, or common partial boundaries, they are candidates for combination, which takes place if the resulting hypothesis has sufficient wall and shadow evidence to support the combined hypothesis. The criterion used for deciding between combining and leaving the hypotheses separate, is whether the confidence associated with the wall and shadow evidence of the composite is greater or less when compared to the sum of the confidence values of the individual hypotheses. This combination is done by deleting the common boundary, and retaining only the non-common boundaries of the two building hypotheses.

Overlap: Two hypotheses may partially overlap. The new hypothesis is obtained by taking the union of the areas of the hypotheses being combined. The combined hypothesis is verified with wall and shadow evidence, and a decision to accept the combination, or not, is made based on whether the confidence associated with the combination is higher or lower than the sum of the confidences of the individual hypotheses.

4.5 Results and Conclusions

So far this system has been tested on views of two different sites. The first site, a modelboard constructed for the RADIUS project, is characterized by a dense array of buildings, aligned parallel to each other; however, it has no vehicles on roads or in parking lots and contains no vegetation. The system has been used with up to four views of this scene in the experiments.

The second site is that of a military base in Fort Hood, Texas; a site more challenging than the modelboard, because it has non-rectangular buildings, vehicles are present on the roads and parking lots, and it has trees and grassy areas. Real lighting conditions cause shadows that are not necessarily the darkest areas in the images. Furthermore the acquisition geometry is such that the epipolar lines between many pairs of views are almost parallel (within 5^{o}) to one of the sides of the buildings (in at least one view) at the site. This causes height estimates to be less reliable and the selection process less certain.

4.5.1 Results on Small Areas

A number of examples of results on small areas have been included. Each of the examples illustrates some observed characteristic or some problem that the system solves or may need to solve. In the example in Figure 4.8, there are a number of dark elongated buildings that give rise to a number of detected parallel lines, hence a large number of parallel hypotheses. In the example illustrated in Figure 4.9, the building labeled A is dark, hence one side merges with its shadow in one view. In Figure 4.10, the largest building in the scene, labeled B, has many parallel superstructures that need to be considered, along with the final hypothesis. The figures indicate good performance under fairly difficult conditions.

Figures 4.11 and 4.12 provide instances of repeated application of the combination routine. Figures 4.11(a) and 4.12(a) show verified rectangular building fragment hypotheses. Figures 4.11(b) and 4.12(b) show the results after combination. Rectangular hypotheses are formed as a result of markings present on the rooftops of the buildings. Three rectangular hypotheses must be combined to form each building in this example.

Figure 4.13 illustrates some failures of the system. The building labeled **C** in both views includes part of the ground on the left side, caused by the detection of a small protrusion from the building on that side. This problem may be solved by a more detailed analysis of the regions; this may require use of more sensitive feature detectors or other region analysis. The building labeled **D** in both views excludes a very narrow part of the building on the right, in the left view. This is a result of a combination of inexact location of features (owing to errors in the location of the detected edges), and errors in the camera models, which lead to errors in matching. Better feature location would enable higher-confidence matching starting from the lowest level in the hierarchy of features. This would eliminate many incorrect matches, and allow the use of tighter tolerances in matching.

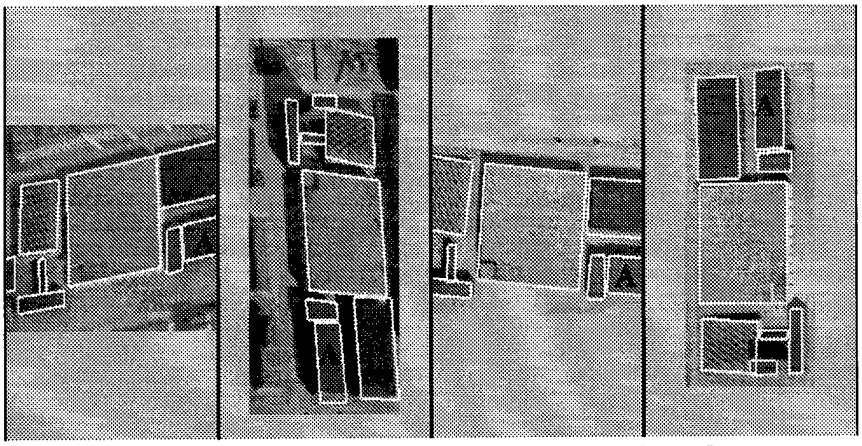


Figure 4.9 Results on a section of the modelboard

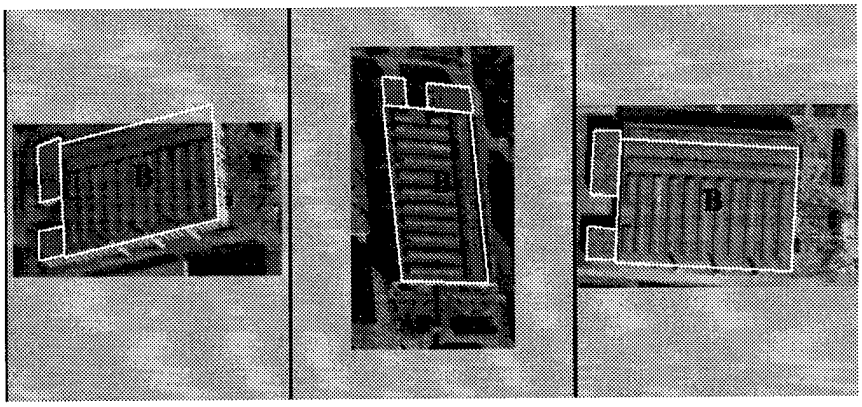
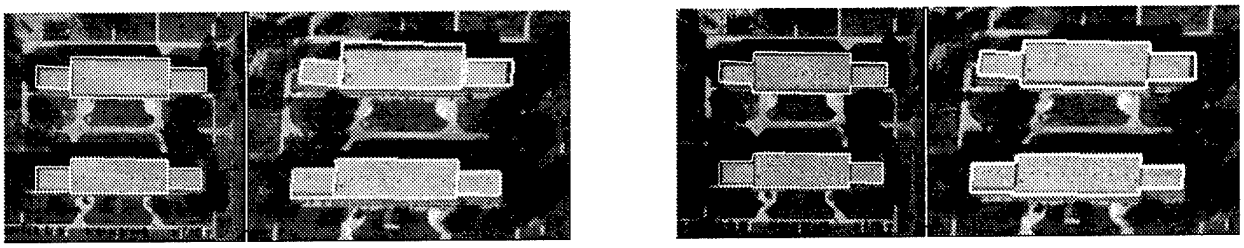


Figure 4.10 Results on another section of the modelboard

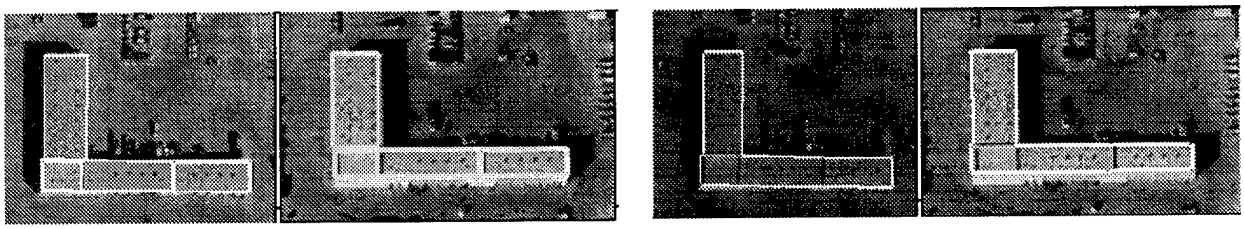
4.5.2 Composite Results

Results obtained for large areas (by processing smaller overlapping areas) of the Fort Hood images are shown in Figures 4.14 and 4.15. These results were obtained by using the depicted view with one other



(a) Rectangular hypotheses (b) Combined Hypotheses

Figure 4.11 Results on a section of Fort Hood



(a) Rectangular Hypotheses (b) Combined Hypotheses

Figure 4.12 Results on another section of Fort Hood

overlapping view. By selecting a window in the depicted view, the system automatically presents the user with a choice of other overlapping views. The composite results were constructed by incrementally augmenting the 3-D model of the site by adding the building models produced from the results of running the system on a selected window. The windows typically accommodated between five and seven buildings. This was necessary because of the memory required for processing larger windows. The characteristics of the areas shown in Figures 4.14 and 4.15 are fairly similar. There are a number of L-shaped buildings, flanked by smaller rectangular buildings. None of these buildings is taller than 15 meters. The system reliably finds the large buildings in areas where the sides of the buildings are not highly fragmented, owing to the similar reflectance properties of the buildings and the ground near it. It performs less reliably when the epipolar lines are parallel to the sides of the buildings as matching these lines is harder than when the lines form a significant angle with the epipolar lines.

Evaluation of the system is performed using quantitative and qualitative criteria. A model is constructed by hand for comparison. A building is declared detected if its roof area overlaps more than 50 percent of a roof of a building in the supplied model. Quantitative measures of the performance of the system may be defined as follows: if t_p is the number of true positive hypotheses, t_n is the number of true negative hypotheses and f_p is the number of false positive hypotheses, then we define the detection percentage as $t_p/(t_p + t_n)$, and the branching factor as $f_p/(t_p + f_p)$. For the part of the site in Figure 4.14, t_p was 51, t_n was 11, and f_p was 5. For the part of the site in Figure 4.15 t_p was 25, t_n was 7 and f_p was 4. Measures of the number of pixels that are correctly labeled as building and non-building pixels also are useful. They are obtained by comparison with the supplied model. These numbers are useful only when the sample space is of a reasonable size, hence they are provided only for the composite results in Figures 4.14, and 4.15, and tabulated in Table 6:

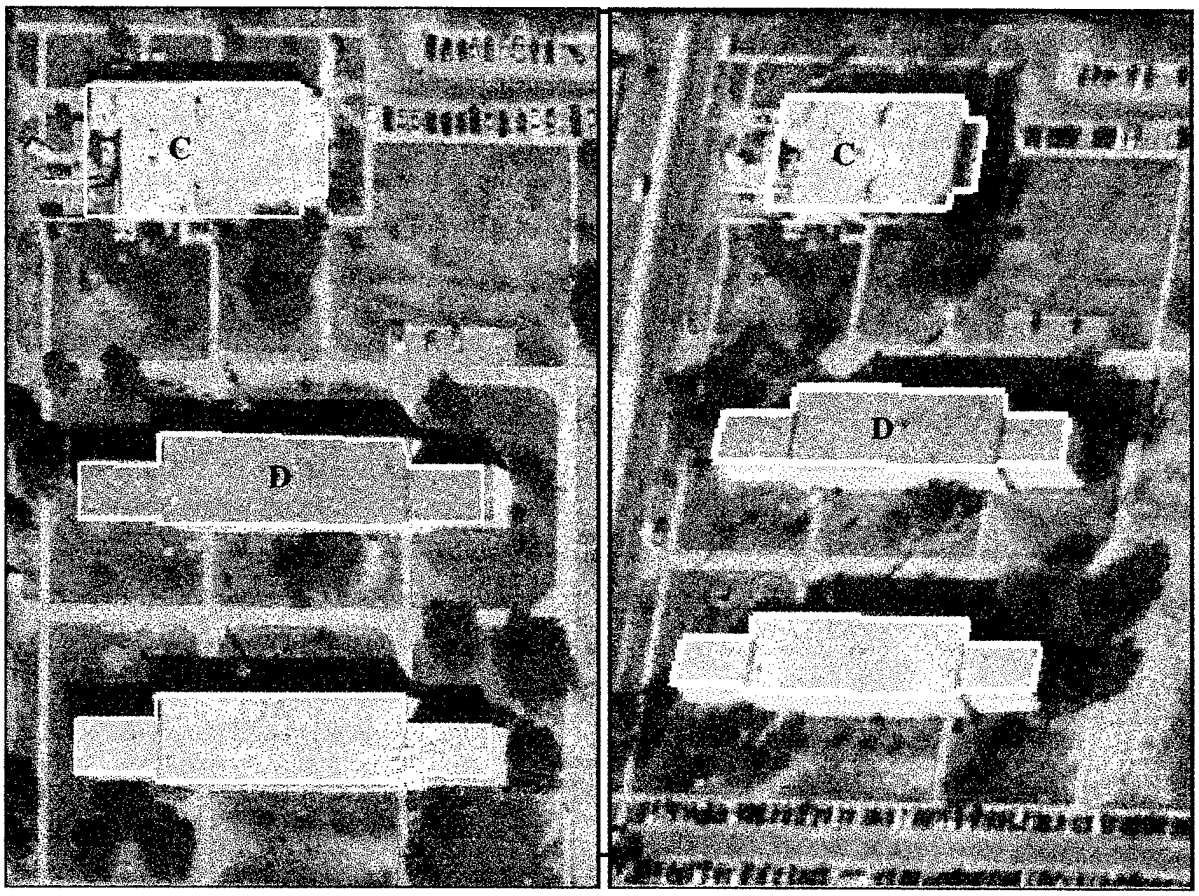


Figure 4.13 Results on a section of Fort Hood (a) on left, (b) on right

Table 6. Performance Evaluation

Figure	Detection Percentage $t_p/(t_p + t_n)$	Branching Factor $f_p/(t_p + f_p)$	Correct building pixels	Correct non-building pixels
Figure 4.14	82.26%	0.08929	75.36%	99.13%
Figure 4.15	78.13%	0.13333	71.84%	98.72%

4.5.3 Conclusions

The system described here has shown some promising results on the problem of automatic model construction. It is able to detect and describe many buildings under widely varying viewpoints and varying times of day. It does not rely on the views being taken at the same time for stereo and trinocular analysis. Currently it is able to detect and describe rectilinear buildings.

Figure 4.14 Results on sections of Fort Hood

Figure 4.15 Results on more sections of Fort Hood

One of the basic problems that remains is the detection and description of buildings that have large occlusions on certain sides due to trees, other buildings, shadows, or other objects. There is a trade-off that occurs in this case. Allowing the system to be very sensitive to small amounts of evidence that might constitute a building hypothesis (a small L-shape that has no obvious match in another view), will result in a large number of hypotheses that the selection mechanism has to decide upon, most of which are not valid. In addition there also will be a much larger number of false alarms. Detection and description of arbitrary polygonal buildings might require the use of line matches alone, because all primitives of greater complexity than lines are specific to rectilinear building detection and description. Detection and description of non-polygonal buildings may require general models for the roofs, such as circular buildings (for storage tanks for example) and hemispherical buildings

5 References

[1] Médioni G., A. Huertas and M. Wilson. *Automatic Registration of Color Separation Films*, Machine Vision and Applications, Springer-Verlag, New York, NY, Vol. 4, 1991, pp 33-51.

[2] Nevatia, R. and A. Huertas. *Research on Model-Based Change Detection and Model Updating*, USC-IRIS Technical Report No. 340, August 1994.

[3] Nevatia, R. and A. Huertas. *Research on Model-Based Change Detection and Model Updating.* USC-IRIS Technical Report No. 350, August 1995.

[4] Huertas, A., M. Bejanin and R. Nevatia. *Model Registration and Validation,* In Automatic Extraction of Man-Made Objects from Aerial and Space Images, Gruen, A., Kuebler, O., Agouris, P., Editors, Birkhauser Verlag, Switzerland, 1995, pp 33-42.

[5] Huertas, A. and R. Nevatia. *Detecting Changes in Aerial Views of Man-Made Structures*, Proceeedings of the ARPA Image Understanding Workshop, Palm Springs, CA, February, 1996, pp 381-388.

[6] Strat, T. et al. *The RADIUS Common Development Environment*, Proceedings of the DARPA Image Understanding Workshop, San Diego, CA, Morgan Kaufman, Publisher, January, 1992, pp 215-226.

[7] Gerson, D. and S. Wood. *RADIUS Phase II. The RADIUS Testbed System,* Proceedings of the Image Understanding Workshop, Vol 1, Monterey, CA, Morgan Kaufman, Publisher, November, 1994, pp 231-237.

[8] Huertas, A., S. Marouani and G. Medioni. *Model-Based Aircraft Recognition in Perpective Aerial Imagery.* IEEE Computer Vision Symposium, Coral Gables, FL, November, 1995, pp 371-376.

[9] Lin C., A. Huertas and R. Nevatia. *Detection of Buildings using Perceptual Grouping and Shadows,* Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, June, 1994, pp 62-69.

[10] Lin, C., A. Huertas and R. Nevatia. *Detecting Buildings from Monocular Images,* in Automatic Extraction of Man-Made Objects from Aerial and Space Images, Gruen, A., Kuebler, O., Agouris, P.. Editors, Birkhauser Verlag, Switzerland, 1995, pp 125-134.

[11] Lin, C. and R Nevatia. *Building Detection and Description from Monocular Aerial Images.* Proceeedings of the ARPA Image Understanding Workshop, Palm Springs, CA, February, 1996, pp 461-468.

[12] Noronha, S. and R. Nevatia. *Detection and Description of Buildings from Multiple Aerial Images,* Proceeedings of the ARPA Image Understanding Workshop, Palm Springs, CA, February, 1996, pp 469-478.

[13] Heuel, S.. and R. Nevatia. *Including Interaction in an Automated Modeling System,* Proceeedings of the ARPA Image Understanding Workshop, Palm Springs, CA, February, 1996, pp 429-434.

[14] Russ, T., et al. *VEIL: Combining Semantic Knowledge with Image Understanding,* Proceedings of the ARPA Image Understanding Workshop, Palm Springs, CA, February, 1996, pp 373-380

[15] Mohan, R. and R. Nevatia. *Using Perceptual Organization to Extract 3-D Structures,* IEEE Transactions on Pattern Analysis and Machine Intelligence, 11(11), November, 1989, pp 1121-1139.

[16] Chung, C. and R. Nevatia. *Recovering Building Structures from Stereo,* IEEE Proceedings of Workshop on Applications of Computer Vision, Palm Springs, CA, December, 1992, pp 64-73.

[17] Nevatia R. and R. Babu. *Linear Feature Extraction and Description,* Computer Vision, Graphics and Image Processing, Vol. 13, 1980, pp 257-269.

[18] Canny, J. *A Computational Approach to Edge Detection,* IEEE transactions on Pattern Analysis and Machine Intelligence 8(6), November, 1986, pp 679-698.

[19] Herman, M. and T. Kanade. *Incremental Reconstruction of 3-D Scenes from Multiple, Complex Images*, Artificial Intelligence, 30(3), December, 1986, pp 289-341.

[20] Huertas, A. and R. Nevatia. *Detecting Buildings in Aerial Images*, Computer Vision, Graphics and Image Processing, 41(2), February, 1988, pp 131-152.1988.

[21] Irving, R. and D. McKeown. *Methods for Exploiting the Relationship Between Buildings and their Shadows in Aerial Imagery,* IEEE Transactions on Systems, Man and Cybernetics, 19(6), November/ December, 1989, pp 1564-1575.

[22] Jaynes, C., F. Stolle and R. Collins. *Task Driven Perceptual Organization for Extraction of Rooftop Polygons*, Proceedings of the ARPA Image Understanding Workshop, Monterey, CA, November, 1994, pp 359-365.

[23] Liow, Y. and T. Pavlidis. *Use of Shadows for Extracting Buildings in Aerial Images*, Computer Vision, Graphics and Image Processing, 49, February, 1990, pp 242-277.

[24] McGlone, J. and J. Shufelt . *Projective and Object Space Geometry for Monocular Building Extraction*, IEEE Proceedings of Computer Vision and Pattern Recognition, Seattle, WA, June, 1994, pp 54-61.

[25] Roux , M. and D. McKeown. *Feature Matching for Building Extraction from Multiple Views*, IEEE Proceedings of Computer Vision and Pattern Recognition, Seattle, WA, June, 1994, pp 46-53.

[26] Venkateswar, V. and R. Chellappa. *A Framework for Interpretation of Aerial Images*, Proceedings of the International Conference on Pattern Recognition, June, 1990, pp 204-206.

[27] Neuenschwander, W., P. Fua, G. Szekely and O. Kubler. *Making Snakes Converge from Minimal Initialization*, Proceedings of the International Conference on Pattern Recognition, Seattle, WA, June, 1994, pp 613-615.

[28] Shufelt. J. and D. McKeown. *Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery*, Computer Vision, Graphics and Image Processing, 57(3), May, 1993, pp 307-330.

[29] Witkin, A. and J. Tenenbaum. *On The Role of Structure in Vision* in Human and Machine Vision, J. Beck, B. Hope and A. Rosenfeld, Eds., Academic Press, New York, NY, 1983, pp 481-543.

[30] Collins, R., Y. Cheng. C. Jaynes, F. Stolle and X. Wang. *Task Driven Perceptual Organization for Extraction of Rooftop Polygons*, Proceedings of International Conference on Computer Vision, 6, June, 1995, pp 888-893.

[31] Lim, H. and T. Binford. *Stereo correspondence: A Hierarchical Approach*, Proceedings of DARPA Image Understanding Workshop, Los Angeles, CA, February, 1987, pp 234-241.